MOLECULAR AND CELLULAR BIOLOGY, July 1995, p. 3904–3916 Vol. 15, No. 7
0270-7306/95/$04.00+0

Homologous DNA Sequences and Cellular Factors Are Implicated in the Control of Glucagon and Insulin Gene Expression

MARTINE CORDIER-BUSSAT, CORINNE MOREL, AND JACQUES PHILIPPE*

Departments of Genetics and Microbiology and of Medicine, Centre Médical Universitaire, 1211 Geneva 4, Switzerland

Received 30 November 1994/Returned for modification 20 January 1995/Accepted 20 April 1995

The glucagon gene is specifically expressed in the alpha cells of pancreatic islets. The promoter of the glucagon gene is responsible for this specificity. Within the promoter, the upstream promoter element G1 is critical to restrict expression to the alpha cells. We define here a composite DNA control element, G4, localized upstream of G1 between nucleotides −100 and −140 which functions as an islet-specific activator in both glucagon- and insulin-producing cells but not in nonislet cells. G4 contains at least three protein binding sites. The most proximal site, E2, is highly homologous to the E1, SMS-UE, and B elements of the rat insulin I, somastatin, and elastase I genes, respectively, and interacts with a pancreas-specific complex; the distal site, E3, represents an E box which is identical to the E boxes of the rat insulin I and II genes and binds to a complex similar or identical to IEF1 which has been implicated in the tissue-specific control of insulin gene expression. These two sites necessitate a third element, the intervening sequence, to activate transcription. We conclude that the first 140 bp of the glucagon gene promoter contains at least two DNA control elements responsible for pancreatic alpha-cell-specific expression: G4, an islet cell-specific element sharing common binding sites with the insulin gene, and G1, which restricts glucagon gene expression to the alpha cells. This double control of specificity might have relevance during islet cell differentiation.

The pancreatic islets of Langerhans are composed of at least four distinct hormone-producing cells, alpha, beta, delta, and PP cells, which produce glucagon, insulin, somatostatin and pancreatic polypeptide, respectively. It has been suggested that, during development, a multipotential endodermal cell gives rise to all islet cell phenotypes (1, 34, 37). Glucagon, a 29-amino-acid peptide involved in the control of glucose homeostasis, is the first hormone to be produced in the embryonic pancreas (1, 33). The glucagon-producing alpha cell may thus constitute an excellent model to study the mechanisms governing differentiation and cell-specific control of gene expression.

Characterization of *cis*-acting and *trans*-acting factors that are involved in gene expression has permitted us to gain insights into the molecular mechanisms of cell differentiation (41, 47). Expression of the glucagon gene is restricted to the alpha cells of the pancreatic islets, to the L cells of the intestine, and to specific brain cells (33). Studies with transgenic mice and glucagon-producing cells transiently transfected with DNA constructs containing various lengths of the rat glucagon gene 5′-flanking sequence linked to the chloramphenicol acetyltransferase (CAT) gene have shown that the most proximal 300 bp was sufficient to direct pancreatic alpha-cell-specific expression (10, 12, 33). Four DNA control elements have been defined within this 300-bp region: a proximal upstream promoter element, G1; two enhancers, G2 and G3; and a cyclic AMP (cAMP)-responsive element (CRE) (9, 21, 35).

We have previously shown that the most proximal 168 bp of the promoter sequence was involved in alpha-cell-specific expression of the glucagon gene in transiently transfected cells (35). More recently, we identified a critical sequence between nucleotides (nt) −52 and −100 corresponding to G1 (26). Several indications, however, suggested that additional transcriptionally important elements were present just upstream of G1 between nt −100 and −140. In previous transfection experiments using the first 350 bp of the rat glucagon gene 5′ flank, we found that a mutant with an internal deletion between nt −114 and −144 was less than half as active as the wild-type promoter while different mutations between nt −136 and −165 did not lead to decreased activity (35, 38). Recent studies have also suggested that the sequence from nt −100 to −110 is highly homologous to the B element of the elastase I gene as well as to the E1 and SMS-UE elements of the rat insulin I and somatostatin genes, respectively (23); the B element, when isolated from the other elastase I gene enhancers, has been shown to direct selective expression to a reporter gene in islet cells of transgenic mice (23). Lastly, two E-box motifs which are conserved between rats and humans (35) are found from nt −130 to −135 and nt −103 to −108, respectively; a third E-box motif is also present within G1 between nt −78 and −83. E boxes, whose consensus motif is CANNTG, are known to interact with a specific family of transcription factors characterized by a helix-loop-helix (HLH) DNA-binding motif (27, 28). These factors bind DNA as homo- or heterodimers and are important for the cell-specific expression of immunoglobulin and muscle-specific genes as well as hormone genes expressed in the pituitary and the pancreatic beta cells (18, 20, 30, 46, 49–51). Two of the E boxes present in the glucagon gene promoter are identical to the rat insulin I gene FAR and NIR boxes and to the rat insulin II gene ICE/RIPE3a sequence, which have been demonstrated to interact with the IEF1 complex containing the ubiquitous HLH proteins E12 and E47 and an islet cell-specific factor and to play a key role in the beta-cell-specific expression of both insulin genes (17, 30).

Since we implicated the first 168 bp in the cell-specific expression of the glucagon gene, we decided to assess the functional importance of the region located between nt −100 and −140 and to define the respective roles of the three potential

* Corresponding author. Mailing address: Dept. of Genetics and Microbiology, Centre Médical Universitaire, 9, avenue de Champel, 1211 Geneva 4, Switzerland.

TABLE 1. Oligonucleotides used in this study

Oligonucleotide	Sequence[a]
E-130	5′ gatcc TGAGGAG **CAGATG** AGCAGAGT a 3′ g ACTCCTC **GTCTAC** TCGTCTCA tctag
E-108	5′ gatcc AAAT **CATTTG** AACA a 3′ g TTTA **GTAAAC** TTGT tctag
E-78	5′ gatcc TTTA **CAGATG** AGAA a 3′ g AAAT **GTCTAC** TCTT tctag
E-108/IS/E-130	5′ gatcc TGAGGAG **CAGATG** AGCAGAGTGGGCGAGTGAAAT **CATTTG** AACA a 3′ g ACTCCTC **GTCTAC** TCGTCTCACCCGCTCACTTTA **GTAAAC** TTGT tctag
E-108-N/IS/E-130	5′ gatcc TGAGGAG **CAGATG** AGCAGAGTGGGCGAGTGAAAT **ATCGAT** AACA A 3′ g ACTCCTC **GTCTAC** TCGTCTCACCCGCTCACTTTA **TAGCTA** TTGT tctag
E-108/IS/E-130-N	5′ gatcc TGAGGAG **AGATCT** AGCAGAGTGGGCGAGTGAAAT **GTAAAC** TTGT tctag g ACTCCTC **TCTAGA** TCGTCTCACCCGCTCACTTTA **GTAAAC** TTGT tctag
E-108/IS-N/E-130	5′ gatcc TGAGGAG **CAGATG** AGCAGAGATCGATAGTGAAT **CATTTG** AACA A 3′ g ACTCCTC **GTCTAC** TCGTCTCTAGCTCTCACTTTA **GTAAAC** TTGT tctag
IEB1	5′ gatcc GC **CATCTG** CC A 3′ g CG **GTAGAC** GG tctag
IS	5′ gatcc AGCAGAGTGGGCGAGTGAAATa g TCGTCTCACCCGCTCACTTTA tctag
B-like element (rat insulin I gene)	5′ gatcc TTTCTGGGAAATGAGGTGa g AAAGACCCTTTACTCCACtctag

[a] Mutated bases are underlined. Boldface type indicates E-box motifs.

E boxes by performing both binding and functional studies. We report here that the two distal E boxes, E2 and E3, bind three specific complexes, A, B, and C. B is ubiquitous, whereas A and C are islet cell-specific and found in both glucagon- and insulin-producing cells; complex A binds to E3 and is similar or identical to IEF1, while complex C, which binds E2, has no specificity for the E-box motif. Neither E2 nor E3 can function by itself as a transcriptional activator; they rather function as a transcriptional unit which also includes an intervening sequence (IS) in both glucagon- and insulin-producing cells but not in nonislet cells. By contrast, the proximal E-box motif, E1, located within G1 does not interact by itself with nuclear proteins from glucagon-producing cells. We conclude that the glucagon gene promoter contains a proximal element which restricts expression to the alpha cells and a more distal unit involved in islet-specific expression. Our data suggest that HLH proteins may be involved not only in the islet cell-specific expression of the insulin gene but also in expression of the glucagon gene.

MATERIALS AND METHODS

Oligonucleotides and plasmids. Oligonucleotides were synthesized on a gene assembler (Pharmacia) by the phosphoramidite method and are listed in Table 1. The IEB1 oligonucleotide was generously provided by M. Walker (Weizmann Institute, Rehovot, Israel). Oligonucleotides containing the wild-type E2, E3, and IS sequences (Table 1) with *Bam*HI-compatible ends were inserted into a *Bam*HI site 5′ of the glucagon minimal promoter (nt +51 to −31) linked to the CAT gene (−31 CAT) (35). The glucagon 5′-flanking sequence (nt +51 to −350) linked to the CAT gene (−350 CAT) (35) was mutated on one, two, or six consecutive nucleotides in E2, E3, and IS sequences by site-directed mutagenesis performed as previously described (26). All constructs were sequenced to confirm identity and orientation by the enzymatic method. Nonspecific oligonucleotides used to compete with protein-DNA complexes in electrophoretic mobility shift assays (EMSAs) represented binding sites for hepatocyte nuclear factor 1 (HNF-1) found in the L-type pyruvate kinase gene (6) (5′-CTAGCTGGT TATACTTTAAGCAGG-3′), for nuclear factor Y (NFY) (5′-ATTTTTCTGAT TGGTTAAAAGT-3′) (8), and for the octamer motif (5′-TTTAGAAATG CAAATTACCCAGGTG-3′) (42).

Cell culture and transfection studies. Islet cells [InR1G9 (44), alphaTC1 (39), betaTC1 (11), and RIN 56A (34)] and nonislet cells (Ltk− mouse fibroblasts, BHK-21 Syrian baby hamster kidney cells, and HepG2 and JEG-3 human hepatoma cells) were cultured in RPMI 1640 containing 5% fetal calf serum and 5% newborn calf serum. Islet cells were transfected in suspension by the DEAE-dextran method (10) (except RIN 56A cells, which were transfected by electroporation at 250 V), and nonislet cells were transfected by the calcium phosphate method (35) with 3 μg of the CAT reporter plasmids and 1 μg of the plasmid pSV2Apap to monitor transfection efficiency. pSV2Apap is a plasmid containing the human placental alkaline phosphatase gene driven by the simian virus 40 long terminal repeat (15). pRSVCAT and poCAT were used as positive and negative controls, respectively (40). Cotransfection experiments were performed with 500 ng of either −350 CAT or two copies of the oligonucleotide E-108/IS/E-130 linked to the glucagon minimal promoter along with pE:Id in 1/1, 1/2, and 1/4 ratios. pE:Id is a plasmid which contains the negative HLH regulator Id cDNA driven by the Moloney murine sarcoma virus long terminal repeat and was generously provided by R. Benezra (Memorial Sloan Kettering Cancer Center, New York, N.Y.) (3). Cell extracts were prepared 48 h after transfection and analyzed for CAT and alkaline phosphatase activities as described elsewhere (38). Protein concentrations were determined with a Bio-Rad protein assay kit.

Cell extracts and EMSAs. Nuclear extracts were prepared by the method of Dignam et al. (7) and Schreiber et al. (43). EMSAs were performed essentially as described elsewhere (30). Oligonucleotides were labeled with the Klenow enzyme and both [α-^{32}P]dCTP and the other three cold nucleosides; they were subsequently purified from polyacrylamide gels.

The standard binding reaction mixture contained the following components in a final volume of 15 μl: 25 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) (pH 7.9), 150 mM KCl, 10% glycerol, 5 mM dithiothreitol, 5,000 cpm of end-labeled double-stranded oligonucleotides, 300 ng of poly(dI-dC), 300 ng of poly(dA-dT), and 2 to 16 μg of crude nuclear extracts. Labeled oligonucleotides were added last, and the mixture was incubated for 20 min at 4°C. Following binding, the mixture was loaded directly onto a 4% polyacrylamide gel (acrylamide/bisacrylamide ratio, 40:0.8) containing 40 mM Tris-glycine (pH 8.5). The gel was electrophoresed at 20 mA for 3 h at 4°C with 40 mM Tris-glycine as the running buffer, dried, and subjected to autoradiography. Antibodies to E12/E47 were incubated for 10 min with nuclear extracts before labeled DNA was added. Antibodies to E12/E47 were a gift from C. Murre (University of California, San Diego) (28) and M. Walker (Weizmann Institute) (2).

Methylation interference assays. For methylation interference experiments, oligonucleotides subcloned into Bluescript (Stratagene, La Jolla, Calif.) were first end labeled with [γ-^{32}P]ATP and polynucleotide kinase on the coding and noncoding strands, respectively, following appropriate restriction enzyme digestion. The labeled double-stranded oligonucleotides were then purified from the plasmid vector on a 6% acrylamide gel, partially methylated by using dimethyl sulfate as described elsewhere (24), and used as probes in binding reactions. The binding reactions were performed in the same way as the EMSAs were, except that they were scaled up eightfold and 200,000 cpm of probe was used. Transfer and elution of protein-DNA complexes were performed as described by Kobr et al. (22), except that 40 mM Tris-glycine buffer was used for the transfer. The bound oligonucleotides were then cleaved with piperidine, and samples were run on 12% polyacrylamide sequencing gels according to the method of Maxam and Gilbert (24).

RESULTS

The glucagon gene promoter contains three E-box motifs. We previously demonstrated that the rat glucagon gene pro-

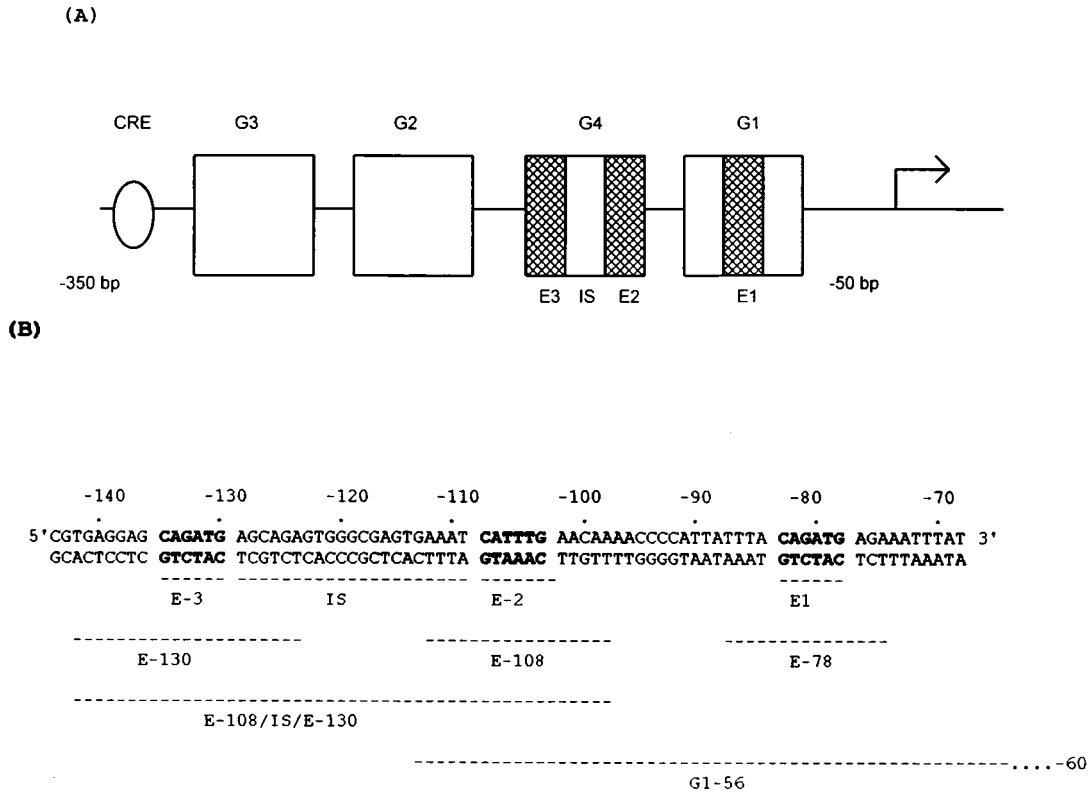

FIG. 1. (A) Schematic representation of the first 350 bp of the rat glucagon gene 5′-flanking region. Boxes and circles represent control elements. (B) DNA sequence of the rat glucagon gene promoter between nt −68 and −144. Fragments used for EMSAs and methylation interference experiments are represented below the sequence. E designates E-box motifs, and CRE indicates a cAMP-responsive element. The arrow in panel A represents the transcriptional start site.

moter (nt +51 to −168) was critical for alpha-cell-specific expression (35). We found that within the promoter, G1, a control element localized between nt −52 and −100, was a major determinant for this specificity (26). Analyses of the glucagon gene promoter sequences revealed the presence of three E-box motifs (Fig. 1 and Table 2); the most proximal E box lies in the core of G1, from nt −78 to −83 (E1), whereas the middle and distal boxes are located from nt −103 to −108 (E2) and nt −130 to −135 (E3), respectively. Previous data suggested that these sequences may be important for basal transcriptional activity; in transfection studies of linker-scanning and internally deleted mutants of the rat glucagon gene 5′-flanking region, we found that disruption of E1 decreased transcription by 90% and that deletion of a 29-bp fragment containing E3 reduced activity by more than 50% (35). By contrast, mutations between nt −136 and −165 did not change activity (35, 38). We thus hypothesized that the three E-box motifs might be functionally important and have relevance in the cell-specific expression of the glucagon gene.

TABLE 2. E-box motifs studied

Gene	Motif	Sequence[a]
Glucagon	E1	TCT **CATCTG** TAA
	E2	AAT **CATTTG** AAC
	E3	GCT **CATCTG** CTC
Insulin I	FAR	CGG **CATCTG** GCC
	NIR	CGC **CATCTG** CCT
Insulin II	ICE	GGC **CATCTG** CTG

[a] Boldface type indicates E-box motifs.

To more precisely evaluate the role of E2 and E3 in transcriptional activity, we first engineered 6-bp mutations within E2, E3, and the IS by site-directed mutagenesis of the first 350 bp of the rat glucagon gene 5′ flank. Plasmids containing mutant and wild-type glucagon gene sequences linked to the CAT gene were then transiently transfected into the hamster glucagon-producing cell line, InR1G9, and their activities were measured. Mutations of either E2 or E3 resulted in a 43% or a 75% decrease in activity, respectively, while only 20% of the wild-type activity was retained with mutations of both sequences (Fig. 2). Interestingly, mutation of IS also reduced transcription by 62%, indicating that the whole region between nt −100 and −140, in contrast to sequences between nt −136 and −165, is critical to basal transcription of the glucagon gene. From these and previous data, we conclude that disruption of any of the three E-box motifs of the glucagon gene promoter affects transcription and that these sequences are essential for promoter activity.

E1, E2, and E3 do not bind the same protein complexes. To investigate whether E1, E2, and E3 interact with DNA-binding proteins, we performed EMSAs using ^{32}P-labeled oligonucleotides containing the different E-box motifs (Fig. 1 and Table 1) and nuclear extracts from InR1G9 cells. As shown in Fig. 3A, with increasing amounts of extracts from 2 to 16 μg, five different complexes bound oligonucleotide E-130 containing E3. Only the upper complex was found to be specific; however, the formation of the four fastest-migrating complexes was inhibited nonspecifically (Fig. 3B and 4A and data not shown). When using the E-108 oligonucleotide corresponding to E2,

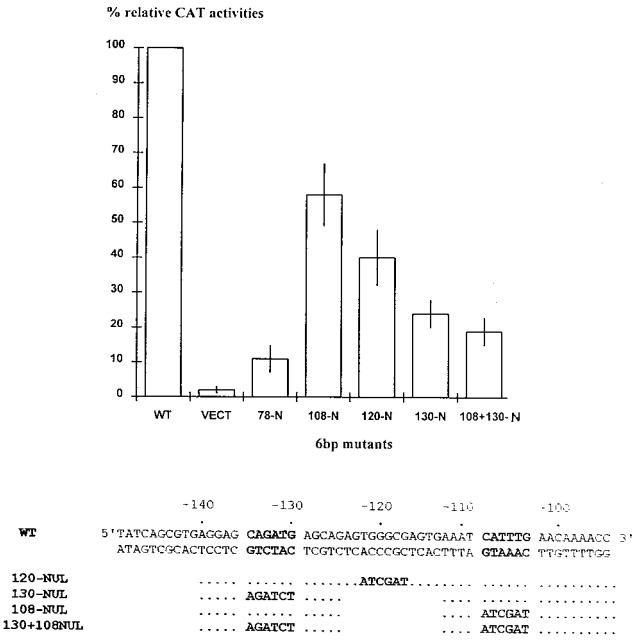

FIG. 2. Mutational analyses of the rat glucagon gene promoter. CAT plasmids containing the wild-type (WT) sequence or a mutated 350-bp stretch of the 5′-flanking region of the glucagon gene were transfected along with a control plasmid, pSV2Apap, into glucagon-producing cells, InR1G9. CAT activities represent CAT/placental alkaline phosphatase ratios and are expressed relative to the value obtained with the WT plasmid. Values are indicated along with the standard errors of the mean (vertical lines) (n = 8). VECT designates the promoterless plasmid poCAT (38). Results obtained with a plasmid mutated in the E1 box (78-N) are presented for comparison (32). WT and mutant sequences are indicated below the diagram.

we detected two complexes, B and C (Fig. 3A), with 2 to 16 μg of extracts. Both complexes migrated faster than the A complex and were specific inasmuch as a large excess of unrelated oligonucleotides was unable to compete (Fig. 3C, lanes 2 to 4).

By contrast to the data obtained with E-130 and E-108, we were unable to detect any specific protein complex binding to E-78 (E1), although the E1-box motif is identical to E3 (data not shown). This was not totally unexpected since we previously characterized three related complexes, B1, B2, and B3, interacting with G1 between nt −76 and −91, including the E1 sequence; these complexes displayed no particular specificity for E1, however (as assessed by point mutations of and around E1), and were not competed for by E-78 (26). Sequences within E1 are thus critical but not sufficient for B1, B2, and B3 binding, and the E1 motif is unable by itself to bind nuclear proteins from glucagon-producing cells.

To further characterize the relative specificities of the A, B, and C complexes for the three different E-box motifs, we performed cross-competition studies. Ten nanograms of unlabeled E-130 was sufficient to efficiently compete for complex A, whereas no competition was obtained with up to 50 ng of E-108 (Fig. 4A). Intermediate effects were observed with E-78, which has an E-box motif identical to that of E-130. Whereas 10 ng of E-78 did not displace complex A, A completely disappeared with 25 and 50 ng (Fig. 4A). Competition for the B and C complexes was most effective with E-108, although competition was also observed with E-130 (Fig. 4B).

Our results indicate that complex A specifically interacts with E-130 and interacts to a much lesser degree with E-78 (as indicated by binding and competition studies) and E-108. Complexes B and C exhibit maximal specificities to E-108 compared with those to E-78 and E-130. These data are consistent with the existence of protein-protein interactions between the different E-108 and E-130 binding factors.

The A, B, and C complexes display different specificities for the E-box motif. Proteins of the HLH family are known to interact with the E-box motif, and their binding is inhibited by mutations within the consensus sequence CANNTG. To more fully characterize the A, B, and C complexes, we tested the consequences of point mutations in and around the E2 and E3 motifs on their binding (Fig. 5A).

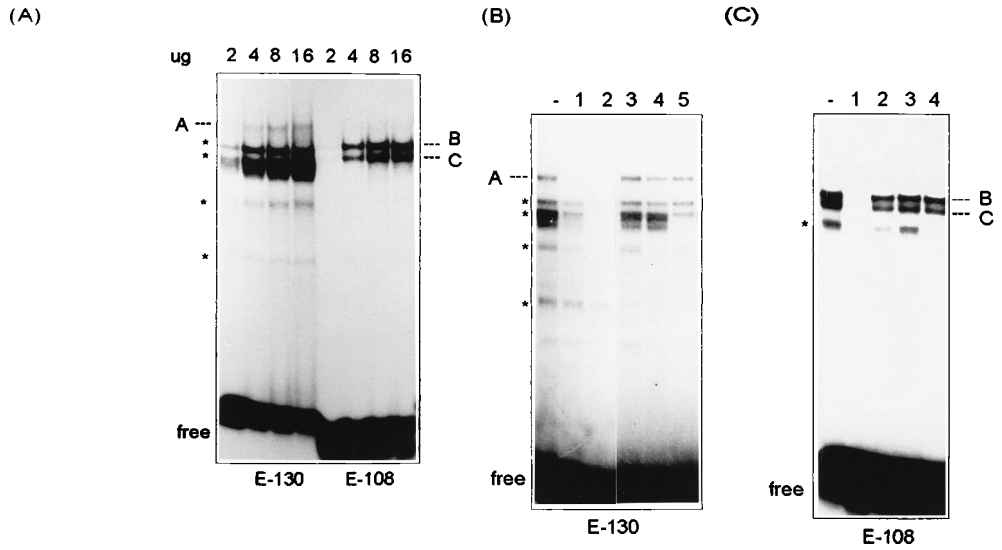

FIG. 3. Binding of nuclear proteins from InR1G9 cells to the E-130 and E-108 oligonucleotides. EMSAs were performed as described in Materials and Methods. (A) ^{32}P-labeled E-130 and E-108 were incubated with increasing amounts of nuclear extracts (from 2 to 16 μg). (B) ^{32}P-labeled E-130 was incubated with 8 μg of InR1G9 nuclear extracts in the absence (−) or presence of the following unlabeled competitor oligonucleotides: 25 and 50 ng of E-130 (lanes 1 and 2, respectively) and 100 ng of three different oligonucleotides unrelated to E-130 (lanes 3 to 5) representing the HNF-1, Oct-1, and NFY binding sites, respectively (see Materials and Methods for sequence). (C) ^{32}P-labeled E-108 was incubated with 8 μg of InR1G9 nuclear extracts in the absence (−) or presence of the following unlabeled competitor oligonucleotides: 50 ng of E-108 (lane 1) and 100 ng of three different oligonucleotides unrelated to E-108 (lanes 2 to 4). A, B, and C indicate the positions of the respective specific protein complexes, while stars indicate nonspecific complexes; free indicates free DNA.

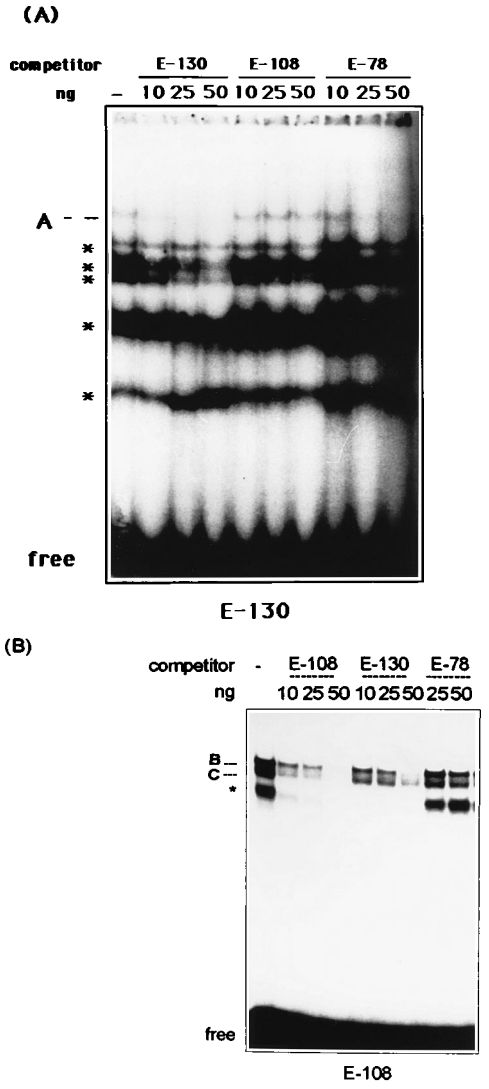

FIG. 4. Competition for the binding of the A, B, and C complexes with oligonucleotides containing the three E-box motifs. EMSAs were performed as for the experiment shown in Fig. 3 with 8 μg of InR1G9 nuclear extracts and ^{32}P-labeled E-130 (A) or E-108 (B) in the absence (−) or presence of the indicated amounts of unlabeled competitor oligonucleotides. A, B, and C indicate the positions of specific protein complexes; stars indicate nonspecific complexes; and free indicates free DNA.

Whereas all mutations of the E3-box motif (M1, M3, and M4) completely abolished binding of complex A, mutations outside the motif (M2 and M5) had no effect (Fig. 5B). Of note, new complexes were observed to bind mutant M1. Correspondingly, effective competition for the A complex was seen with mutants M2 and M5 and not with M1, M3, and M4 (data not shown). The three mutations of the E-box motif of E-108 (V1, V2, and V3) resulted in different consequences for binding of complexes B and C (Fig. 5C). V1 and V3 abolished and V2 severely decreased binding of B; binding of C, by contrast, was not affected by V1 and was moderately affected by V2 and V3. Only slight alterations in binding of B and C resulted from mutations outside the E-box motif (V4 and V5). These results suggest that complex A and possibly complex B, but not complex C, exhibit specificities for the motif CANNTG and thus may belong to the HLH class of DNA-binding proteins.

To investigate the consequences of these mutations on transcriptional activity, we performed site-directed mutagenesis of the first 350 bp of the glucagon gene 5′ flank linked to the CAT gene (Fig. 5A). The resultant plasmids were transiently transfected into InR1G9 cells, and their CAT activities were measured. Whereas mutation of the entire E3-box motif resulted in a 75% decrease in activity, mutations M3 and M4, which abolished binding of complex A in vitro, reduced activity by only 35 and 15%, respectively. Additional double mutants, M6 and M7, which also abolished binding of complex A (data not shown), resulted in more pronounced decreases in activity: up to 40 and 63%, respectively. By contrast, changes of nucleotides outside the motif (M2, M5, and M8) increased transcription by 30 to 70% (Fig. 5D). M8, like M2 and M5, did not alter the binding affinity of complex A for E3 (data not shown). Mutant M1 was not tested since additional specific complexes not detected with the wild-type sequence were observed. The reasons for the relatively modest effects of M3 and M4 on transcription are unclear; it is possible, however, that protein binding to these mutants might be less affected in vivo than in vitro. Similar observations were made previously (20, 36). Alternatively, we cannot formally exclude the possibility that factors other than those of the HLH family regulate E3 activity or that our mutagenesis strategy affects the binding of neighboring factors and thus activity.

All mutations of E2, whether inside or outside the E-box motif, resulted in decreased activity (except V2) (Fig. 5D). There was no correlation, however, between changes in binding of complexes B and C and transcriptional activity. When we assessed the functional consequences of mutations in the E2-box motif (M4) combined with mutations in E3 (V1 to V5), we globally observed an additive effect of the respective single mutations, indicating that E2 and E3 by themselves are unlikely to act synergistically on transcription.

Overexpression of Id-1 decreases transcriptional activity conferred by G4. To further strengthen the hypothesis that E3 binds a transcriptional factor of the HLH class, we cotransfected either −350 CAT or two copies of the oligonucleotide E-108/IS/E-130 representing G4 linked to the minimal glucagon promoter element [2×(E-108/IS/E-130)] along with an expression vector, either alone or containing the negative HLH regulator, Id-1 (3), in different ratios. E-108/IS/E-130 functions as an islet cell-specific enhancer (see Fig. 9). Cotransfection of the expression vector with the Id-1 cDNA resulted in a decrease in the transcriptional activity observed with either −350 CAT or 2×(E-108/IS/E-130) and the expression vector alone (Fig. 5E). Decreases in the ratios of indicator plasmid/expression vector and Id resulted in a dose-dependent attenuation of transcription with quantitatively similar effects whether the indicator plasmid was −350 CAT or 2×(E-108/IS/E-130). These results indicate that sequences within G4 and most likely E3 contain a functional binding site for an HLH binding protein.

The A complex is present in both alpha and beta cells and is similar or identical to IEF1. To investigate the cellular distribution of the A, B, and C complexes, we tested nuclear extracts obtained from different cell lines of islet and nonislet origins. The A complex (Fig. 6A) was observed to be present in both glucagon-producing (InR1G9 and alphaTC1, lanes 1 and 2) and insulin-producing (betaTC1, lane 3) cells but not in nonislet cell lines (lanes 4 to 6). Similarly, complex C was found only in pancreatic endocrine cells (Fig. 6B, lanes 1 to 3) whereas complex B was ubiquitous (lanes 4 to 6). Of note, a specific complex migrating slightly faster than complex C was present in nonpancreatic endocrine cells. Our results indicate that complexes A and C are found preferentially in islet cells.

The observations that the E3 motif is identical to the FAR

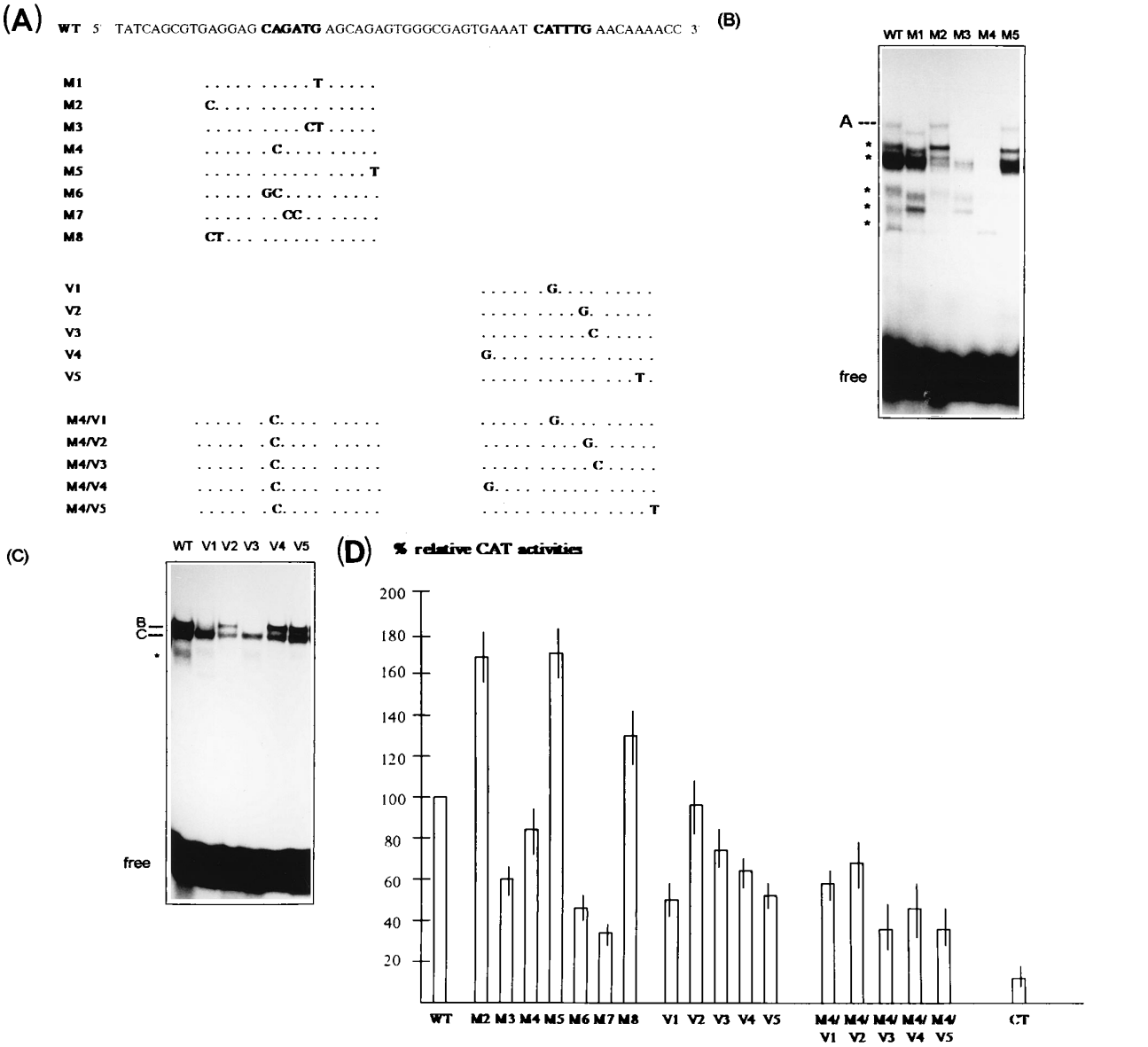

FIG. 5. Binding of nuclear proteins from InR1G9 cells to mutant E-box motifs and functional analyses. (A) Wild-type (WT) and mutant sequences of the glucagon gene promoter. Mutant oligonucleotides used in the EMSAs were of the same size as E-130 or E-108 with the indicated mutations (M1 to M8, V1 to V5, and M4/V1 to V5). Mutations within the 350 bp of the rat glucagon gene 5′-flanking sequence were constructed by site-directed mutagenesis (see Materials and Methods). (B and C) EMSAs were performed as for the experiments shown in Fig. 3 with 8 μg of proteins from InR1G9 nuclear extracts and ^{32}P-labeled wild-type (WT) or mutant E-130 (B) or WT or mutant E-108 (C). The respective labeled oligonucleotides are indicated above the lanes. A, B, and C indicate the respective specific protein complexes; stars indicate nonspecific complexes; and free indicates free DNA. (D) Functional analyses of E-box mutants by transient transfections in InR1G9 cells. Plasmid constructs containing the CAT gene under the control of wild-type (WT) or mutant sequences of the rat glucagon gene were tested for CAT activities after transfection into InR1G9 cells. Each construct was cotransfected with the plasmid pSV2Apap to correct for differences in transfection efficiencies. Relative CAT activities represent CAT/placental alkaline phosphatase enzymatic ratios relative to the value for the WT construct, and the standard errors of the mean (n = 8) are indicated by vertical lines. CT designates the promoterless plasmid, poCAT (38). (E) Effect of overexpression of Id-1 on glucagon gene expression. Plasmid constructs containing the CAT gene under the control of the first 350 bp of the rat glucagon gene 5′ flank (−350 CAT) or two copies of the oligonucleotide E-108/IS/E-130 linked to the minimal glucagon gene promoter [2×(108/IS/130)] were cotransfected with either the expression vector pE (white bars) or the expression vector containing the Id-1 cDNA, pE:Id (black bars), into InR1G9 cells. A 500-ng amount of indicator plasmid was used for transfection along with increasing amounts of either pE or pE:Id (from 500 ng to 2 μg). CAT activities were measured and corrected as described in the legend for panel D and are expressed relative to the activity measured for the indicator plasmid cotransfected with pE in a 1/1 ratio (n = 3). *, $P < 0.05$; ○, $P < 0.01$. Error bars indicate standard errors of the mean.

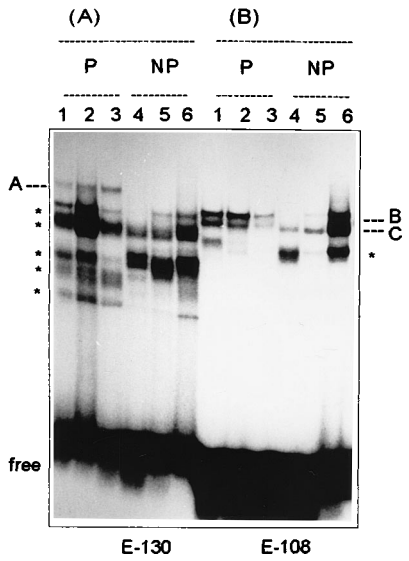

FIG. 6. Cell type distribution of A, B, and C complexes. EMSAs were performed as for the experiment shown in Fig. 3 with 8 μg of cell nuclear extracts and ^{32}P-labeled E-130 (A) or E-108 (B) oligonucleotides. P indicates pancreatic endocrine cells (InR1G9 in lanes 1, alphaTC1 in lanes 2, and betaTC1 in lanes 3), and NP indicates nonpancreatic cells (the human chronic lymphocytic leukemia cell line Cohen in lanes 4, JEG-3 in lanes 5, and HepG2 in lanes 6). A, B, and C indicate the respective specific protein complexes; stars indicate nonspecific complexes; and free indicates free DNA.

and NIR boxes of the rat insulin I gene and to the ICE/RIPE3a element of the rat insulin II gene (Table 2) and that complex IEF1 which interacts with these elements has the same cellular distribution as the A complex has (31) led us to investigate whether IEF1 and complex A are identical. We first used an oligonucleotide, IEB1, containing the NIR box sequence of the rat insulin I gene (references 2 and 48 and Tables 1 and 2) in an attempt to inhibit complex A binding to E-130. IEB1 efficiently competed for binding by complex A, displaying an even better affinity for A than E-130 did (Fig. 7A). With ^{32}P-labeled IEB1 and the same amounts of InR1G9 nuclear extracts, a more intense complex than that observed with E-130 was systematically seen. This complex was competed for by both E-130 and IEB1, more effectively so by the latter binding site, suggesting a better affinity of complex A for IEB1 (Fig. 7B). The nature of this complex was further investigated by adding specific antibodies to the ubiquitous HLH protein E12/E47, previously shown to be contained within IEF1 (2, 29, 48), to the EMSA binding reaction mixture; antibodies from two different origins (see Materials and Methods) inhibited complex binding and induced a supershift (Fig. 7B). The complex binding to IEB1 that we detected thus has the same characteristics as IEF1 does. To study the relationship between IEF1 and complex A, we compared their EMSA migration patterns. As shown in Fig. 7B (rightmost lane), the two complexes had identical patterns. We then assessed whether anti-E12/E47 antibodies would react against the A complex; we observed an inhibition of A-complex formation with antibodies that was not present with preimmune serum (Fig. 7C). We conclude that IEF1 and A are similar or identical and that they both contain the ubiquitous HLH protein E12/E47 or a highly related protein.

The B and C complexes bind E2 and the B-like element of the rat insulin I gene equally well. The B element of the rat elastase I gene was previously shown to be highly homologous to sequences of the E1 element of the rat insulin I gene and to E2 (Table 3) and to direct selective expression of a transgene to islet cells (23). Furthermore, the B element of the elastase I gene binds two complexes that also interact with the E1 sequences of the rat insulin I gene and with E2. We thus investigated whether complexes B and C were competed for by the B-like element of the rat insulin I gene found within E1. Figure 8 shows that the B-like element competes for B and C binding with efficiencies similar to those of E-108. By contrast, the B-like element was unable to compete for complex A even at a 500-fold molar excess (data not shown). These results indicate that E2 and the B-like element of the rat insulin I gene are able to bind the same complexes with similar affinities.

E2, E3, and IS form an islet cell-specific transcriptional unit. To study whether E2 and E3 can activate transcription independently, we multimerized E-108 and E-130 oligonucleotides and subcloned the resulting multimers 5′ upstream of a minimal glucagon gene promoter containing the TATA box (nt +51 to −31). The plasmids were then transiently transfected into glucagon-producing cells (InR1G9, alphaTC1, and RIN 56A). Neither E-108 nor E-130 in up to five copies (Fig. 9) was able to activate transcription. Similarly, two copies of E-108/E-130 sequentially linked and subcloned 5′ of the glucagon gene minimal promoter were inactive. We then tested the activity conferred by two copies of a 44-bp oligonucleotide containing E2, E3, and IS (E-108/IS/E-130); by contrast to E-108/E-130, E-108/IS/E-130 was able to activate transcription by three- to fivefold in alphaTC1, InR1G9, and RIN 56A cells (Fig. 9). These data indicate that E2 and E3 can function neither alone nor together as enhancers; the presence of IS is necessary for the activation of transcription.

Since complexes which interact with E-108 and E-130 are present in both glucagon- and insulin-producing cells, we tested the activities of the different constructs in betaTC1 cells. As expected, plasmids containing multimers of E-108, E-130, or E-108/E-130 did not activate transcription above the basal level while E-108/IS/E-130 increased activity by 5.5-fold in betaTC1 cells, a result quantitatively similar to that obtained with glucagon-producing cells. None of the plasmids could activate transcription in nonislet cell lines (Fig. 9). We conclude that E-108/IS/E-130 (nt −100 to −140) functions as an islet-specific activating element that we name G4.

The IS interacts with islet cell-specific complexes. The IS must be important for glucagon gene expression since its mutation results in a 62% decrease in activity and its presence is necessary for the E-108/IS/E-130 functional unit to activate transcription. We investigated whether IS was a specific binding site for nuclear proteins from InR1G9 cells. E-108/IS/E-130 was used as the binding site in EMSAs at protein concentrations at which complexes binding to E-130 (Fig. 10A) or E-108 were easily detected. At low protein concentrations, we observed two complexes, D and E (Fig. 10A). As protein concentrations increased, E progressively became predominant and D nearly disappeared. Both complexes were specific and different from the A, B, and C complexes. Competition for complex D was achieved only by oligonucleotide E-108/IS/E-130 (Fig. 10B, lane 2, 2 μg of protein) and not by E-108, E-130, or the octamer binding site (lanes 3 to 5), indicating that complex D interacts with IS. Complex E was efficiently competed for by E-108/IS/E-130 (Fig. 10B, lane 2, 16 μg of protein) and also, to a lesser extent, by E-108 and E-130 (lanes 3 and 4) but not by the octamer binding site (lane 5). Addition of E-108 and E-130 decreased binding of complex E by 52 and 66%, respectively, while markedly increasing complex D, as assessed by laser densitometry. These results suggest that while complex D contacts only IS, complex E requires E-108 and E-130 to form.

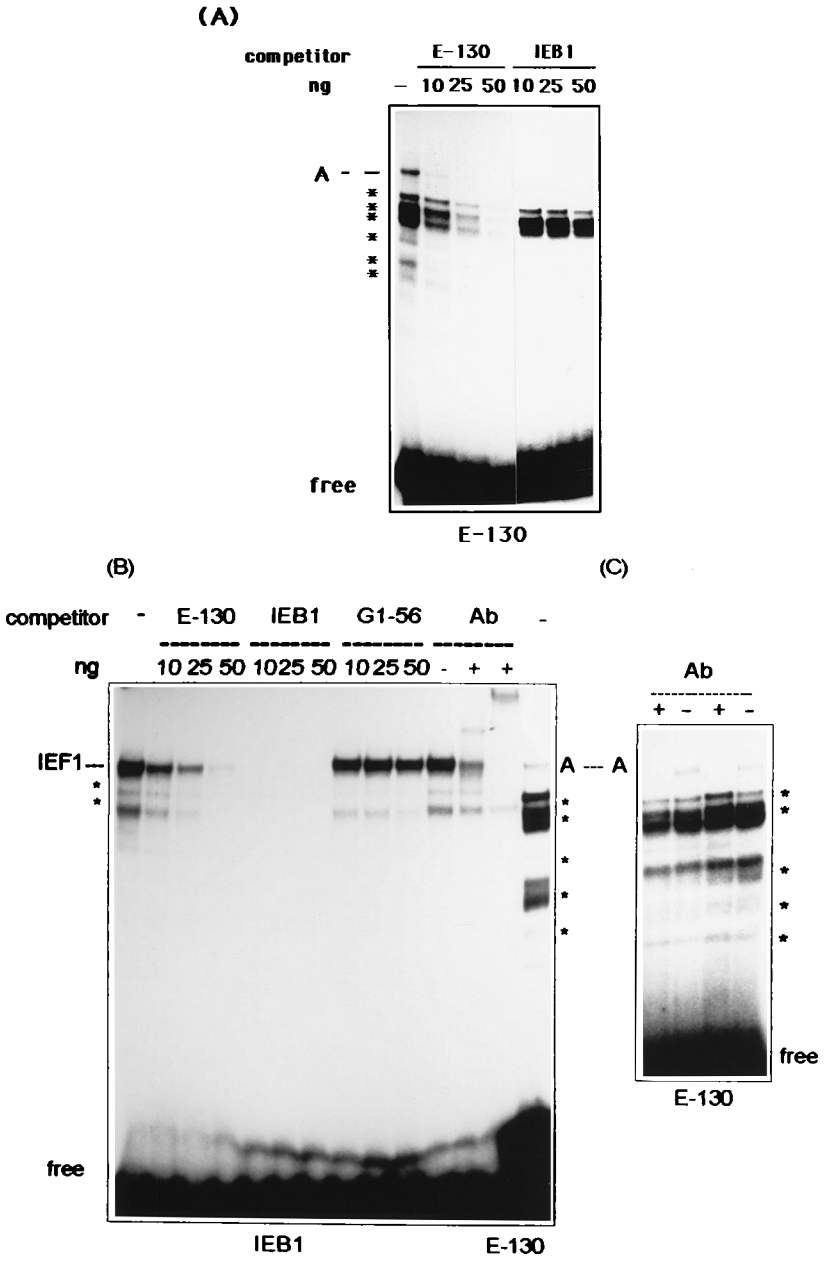

FIG. 7. The A and IEF1 complexes are similar or identical. EMSAs were performed with 8 μg of InR1G9 nuclear extracts and ^{32}P-labeled E-130 (A, B, and C) and IEB1 (B) oligonucleotides. (A) Competition for the A complex by E-130 or IEB1 at the amounts indicated above the lanes. (B) Competition for IEF1 by E-130, IEB1, or nonspecific DNA (G1-56) at the amounts indicated above the lanes. Ab indicates the addition of preimmune antiserum (−) or of two different antibodies against E12/E47 (+) (2, 25) to the incubation reaction mixture. The rightmost lane illustrates the migration pattern of the A complex bound to E-130 for comparison. (C) Preimmune serum (−) or two different antibodies against E12/E47 (+) (2, 25) were added to the incubation reaction mixture containing 8 μg of InR1G9 nuclear extracts and ^{32}P-labeled E-130. IEF1 and A indicate the respective specific protein complexes, stars indicate nonspecific complexes, and free indicates free DNA.

To further assess the involvement of E2 and E3 in the formation of the E complex, we performed EMSAs with wild-type and mutant E-108/IS/E-130 oligonucleotides at protein concentrations allowing detection of complexes A, B, and C (Fig. 10C); mutations were within either E2 (E-108N), E3 (E-130N), or IS (IS-N). As shown in Fig. 10C, neither E-108N nor E-130N affected complex D (at 2 μg of extracts) while both decreased complex E binding, by 41 and 73%, respectively (with a corresponding increase in complex D). Mutation of the IS had the most dramatic effects: both complexes D and E disappeared. We then compared the patterns of complexes binding to E-108/IS/E-130 and to IS alone. As shown in Fig. 10D, complex D and an additional complex (F) were found to bind IS but complex E did not bind. Complex F was also faintly visible with E-108/IS/E-130, but complex E formation was vastly favored as protein concentration increased. Overall, these data suggest that complex E contains complex D and additional factors which require E2 and E3 for maximal stability.

To more precisely define the binding site of complexes D and E, we performed methylation interference assays (Fig. 11A). We obtained the same pattern of decreased methylated

TABLE 3. B-like elements

Gene	Element	Sequence
Glucagon	E2	TCAAATGATTT
Insulin I	E1	GGAAATGAGGT
Somatostatin	UE	TTGATTGATTT
Elastase I	B	ATAAATGAGTT

G's for both complexes. Affected G's were localized to nt −117, −119, −120, −121, and −123 on the upper strand and to nt −118 on the lower strand, establishing the minimal binding site for D and E as GTGGGCG (Fig. 11B). The fact that we observed the same pattern for D and E is consistent with the view that E contains D; we were not able, however, to detect additional affected G's in the E boxes when we analyzed either complex E or complexes A, B, and C.

We next studied the cell type distribution of the D and E complexes (Fig. 12). D and E were both present in glucagon-producing (alphaTC1) and insulin-producing (betaTC1) cells, whereas these complexes were not detected in the nonislet cell lines tested. We conclude that E-108/IS/E-130 functions as an islet-specific activating sequence and interacts with islet-specific protein complexes.

DISCUSSION

Determining the functional role of the E-box motifs present in the first 168 bp of the glucagon gene promoter, which had been demonstrated to be important for cell-specific expression, was of major interest. Indeed, the E-box motif CANNTG interacts with one class of transcription factors, the HLH protein family. Members of this family have been shown to be involved in cell type specification through their activation of specific genes such as, for example, the MyoD gene family in muscle cells and the daughterless gene in *Drosophila* neuronal cells (5, 51). E boxes have also been identified in the promoters of several genes transiently or permanently expressed in the beta cells of the pancreatic islets (30, 49, 50). Beta-cell-specific expression of the rat insulin I and II genes appears to be critically dependent on the integrity of their respective E boxes (17, 20, 30).

Our results demonstrate a functional role for a new control element located upstream of G1 between nt −100 and −140. This element contains at least three binding sites: two E-box motifs, E2 and E3, and an IS. E2 interacts with an islet cell-specific C complex which shows no specificity for the E-box motif and a ubiquitous B complex; our mutational analyses suggest that the latter complex may have specificity for an E-box motif, but they are too limited for more definitive conclusions. The lack of correlations between impairment of binding of complexes B and C and alterations in transcriptional activity raised some doubts about the functional relevance of these complexes. However, mutation of the bases surrounding the E2 motif (V4 and V5) which affect transcription but not B and C binding may result in decreased stability of other protein complexes. V4 is indeed located in the IS, and V5 is located at the 5′ end of the G1 element.

E2 is homologous to other DNA elements, E1, SMS-UE, and B found in the rat insulin I, somatostatin, and elastase I genes, respectively (reference 22 and Table 3). The C complex is likely to correspond to BTF1, a factor previously described to bind to the B element of the elastase I gene enhancer (23). The B element (TAAATGAGTTG) binds two complexes; BTF1 is found only in pancreatic endocrine cells, while the

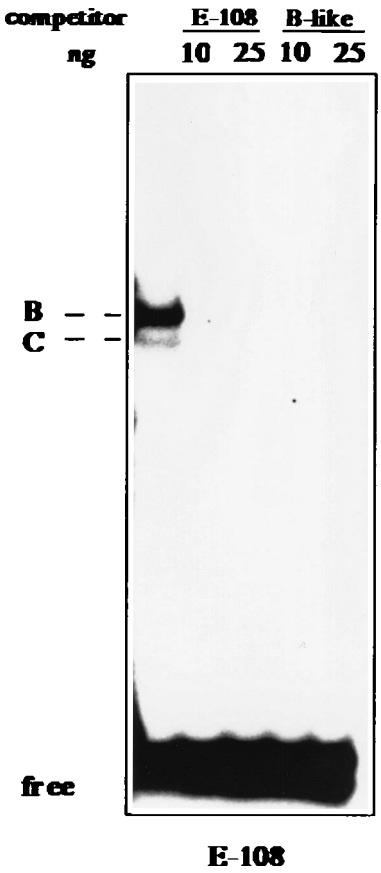

FIG. 8. Complexes B and C bind to both E2 and the B-like element of the rat insulin I gene. EMSAs were performed with 8 μg of InR1G9 nuclear extracts and ^{32}P-labeled E-108. Competition for the B and C complexes was with 10 and 25 ng of unlabeled E-108 or B-like element of the insulin I gene. B and C indicate the positions of specific protein complexes, and free indicates free DNA.

second one is ubiquitous. Using an oligonucleotide containing E-108 and nuclear extracts from insulin-producing cells, Kruse et al. detected two complexes which comigrated in EMSAs with complexes binding to the B element (23). These two complexes also interacted with control elements of the rat insulin I (EI) and the somatostatin (SMS-UE) genes. These results are in agreement with our data showing that E-108 and the B-like element of the rat insulin I gene bind the B and C complexes with similar affinities. We thus propose that the two complexes binding to the B element correspond to the B and C complexes of E-108 and that BTF1 is not beta cell specific but is present in both alpha and beta cells. By this hypothesis, the B complex cannot contain HLH proteins, inasmuch as the E-box motif is disrupted in the B element.

The B element of the elastase I gene performs a binary function. Within the context of the elastase I enhancer, the B element contributes to appropriate acinar cell expression. When separated from other enhancer components, however, the B element selectively directs transcription of a reporter gene to islet cells of transgenic animals. The presence of potential B-element equivalents in other pancreatic genes, both endocrine and exocrine specific, suggests that the B element and its associated factor, BTF1, may play a role in pancreatic gene expression, in particular, as proposed by Kruse et al. (23), in the early stages of pancreatic development when endocrine and exocrine compartments share a common precursor lin-

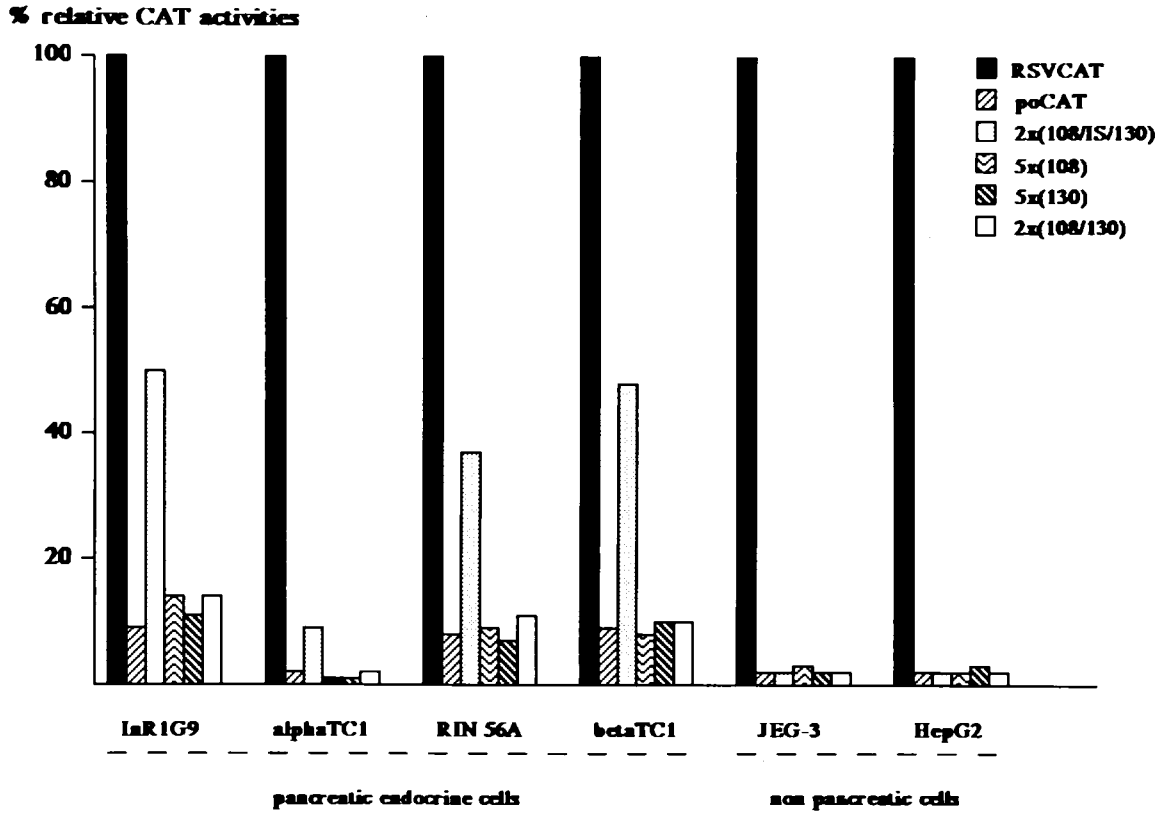

FIG. 9. E-108/IS/E-130 functions as an islet cell-specific enhancer. Plasmids containing five copies of multimerized E-108 or E-130 oligonucleotides, two copies of E-108/E-130 oligonucleotides in tandem [2×(108/130)], or two copies of E-108/IS/E-130 [2×(108/IS/130)] oligonucleotides subcloned 5′ of a minimal glucagon gene promoter (nt +51 to −31) were transiently transfected in glucagon-producing (InR1G9 and alphaTC1) and insulin-producing (betaTC1) cells and in nonislet cells (JEG-3 and HepG2). CAT activities represent CAT/placental alkaline phosphatase enzymatic activity ratios and are expressed relative to the value obtained with the positive control, RSVCAT (38). poCAT (the CAT gene without the promoter) was used as a negative control.

eage. Differentiation of the endocrine lineage could then involve the continued presence of BTF1 and the activation of additional transcription factors specific for the pancreatic hormone genes. Involvement of E2 in the control of glucagon gene expression strengthens this hypothesis.

Analyses directed at E3 indicate that this is the only sequence of the glucagon gene promoter which is likely to function as an E-box motif. Overexpression of Id-1 decreases transcriptional activity conferred by G4 and strengthens this hypothesis. E3 is identical to the FAR and NIR boxes of the rat insulin I gene and to the ICE/RIPE3a sequence of the insulin II gene. E3 interacts with complex A, which is similar or identical to IEF1. IEF1 is an islet cell-specific factor which has been proposed to be involved in the beta-cell-specific expression of the rat insulin genes as well as the gastrin and the secretin genes (17, 20, 29, 30, 33, 49, 50). It is a heterodimeric complex composed of E12/E47 and an unidentified protein present in alpha and beta cells (31, 32). Complex A indeed comigrates with IEF1 in EMSAs and binds to the IEF1 binding site. In addition, like IEF1, A is islet specific and interacts with antibodies against E12/E47. We thus conclude that the same or similar HLH proteins may participate in the control of both glucagon and insulin gene expression.

E2 and E3 have no activating potential by themselves; they must be linked to the IS to activate transcription. The functional role of IS was first suggested by the 62% loss of transcriptional activity which resulted from its mutation. In addition, this mutation impaired the in vitro binding of two pancreas-specific protein complexes, D and E. The minimal recognition site of D and E is GTGGGCG, as assessed by methylation interference assays; D is likely contained within complex E, together with other proteins. E2 and E3 also appear to be necessary for maximal stability of the E complex; whether E contains the A, B, and C complexes remains unclear at present. The D recognition site has some similarities to Sp1 binding sites (19); the cellular distribution of the D complex suggests, however, that Sp1 and D are different. Interestingly, GC-rich sites have been identified just 3′ of the E-box motif of the secretin and gastrin gene promoters (25, 49); this may indicate that cooperativity between HLH proteins and other complexes may be necessary for hormone genes to be expressed in the islets. We have not found, however, DNA sequences homologous to the D/E binding sites within the insulin gene promoters. It thus remains to be seen whether the D/E complexes can interact with insulin genes. Thus, although homologous DNA sequences and cellular factors may be involved in the control of insulin and glucagon gene expression, the functional relationships between E2 and E3 of the glucagon gene and the NIR and FAR boxes and E1 of the rat insulin I gene are clearly different. E2 and E3 form a functional unit with IS, whereas E1 and the NIR and FAR boxes appear to function independently (33). The FAR and NIR boxes of the rat insulin I and II genes, respectively, are part, however, of a functional unit containing multiple binding sites which are critical for beta-cell-specific activity. This arrangement again suggests that E boxes found in islet hormone genes are often

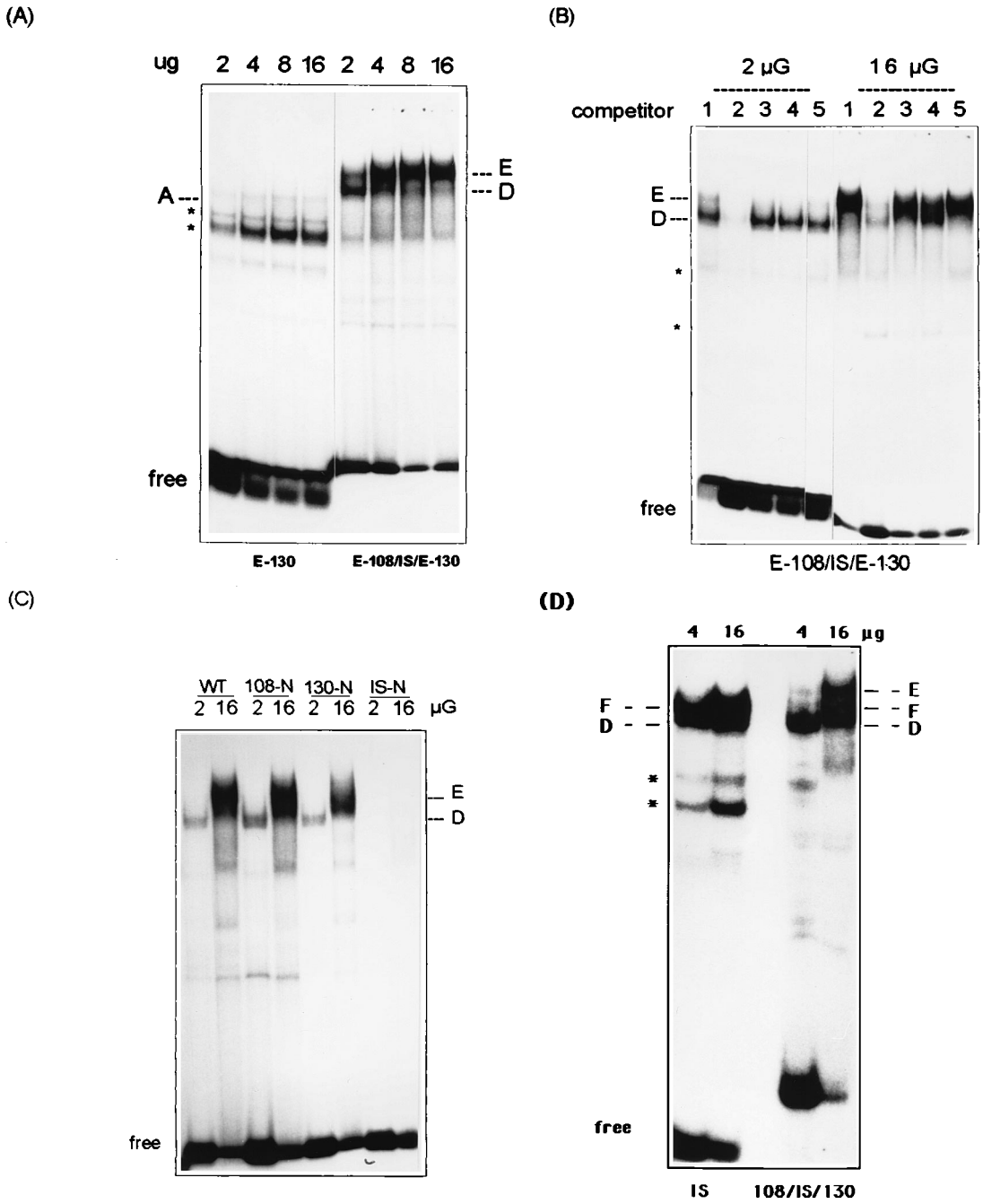

FIG. 10. Binding of InR1G9 nuclear proteins to the IS. EMSAs were performed with ^{32}P-labeled E-130 (A), E-108/IS/E-130 (A, B, C, and D), mutant E-108/IS/E-130 (C), and IS (D) oligonucleotides. (A) Increasing amounts of InR1G9 nuclear extracts (indicated above the lanes) were added to the incubation reaction mixtures. (B) Experiments for competition for the D and E complexes were performed with 2 and 16 μg of InR1G9 extracts by adding 50 ng of the following unlabeled oligonucleotides: E-108/IS/E-130 (lanes 2), E-108 (lanes 3), E-130 (lanes 4), and nonspecific oligonucleotide (lanes 5). Lanes 1 contain no competitor. (C) ^{32}P-labeled wild-type (WT) and mutant E-108/IS/E-130 oligonucleotides (Table 1) were incubated with 2 and 16 μg of InR1G9 nuclear extracts. 108-N, 130-N, and IS-N are abbreviations for E-108-N/IS/E-130, E-108/IS/E-130-N, and E-108/IS-N/E-130, respectively (Table 1). (D) Complex formation on IS and E-108/IS/E-130 oligonucleotides (Table 1). ^{32}P-labeled IS and E-108/IS/E-130 (108/IS/130) were incubated with 8 and 16 μg of InR1G9 nuclear extracts, respectively. A, D, E, and F indicate the respective specific complexes; free indicates free DNA; and stars indicate nonspecific complexes.

part of a functional unit and that protein-protein interactions between HLH and other nuclear proteins may be important for islet cell-specific expression.

The E-108/IS/E-130 functional unit (G4) confers islet-specific activation to the glucagon gene. This is in agreement with the fact that complexes which interact with G4, with the exception of B, are only or preferentially found in islet cells. G4 thus confers to the glucagon gene a level of tissue-specific expression different from that conferred by G1, which restricts expression to the alpha cells (26).

The most proximal E box, E1, present within the upstream promoter element, G1, does not interact by itself with nuclear

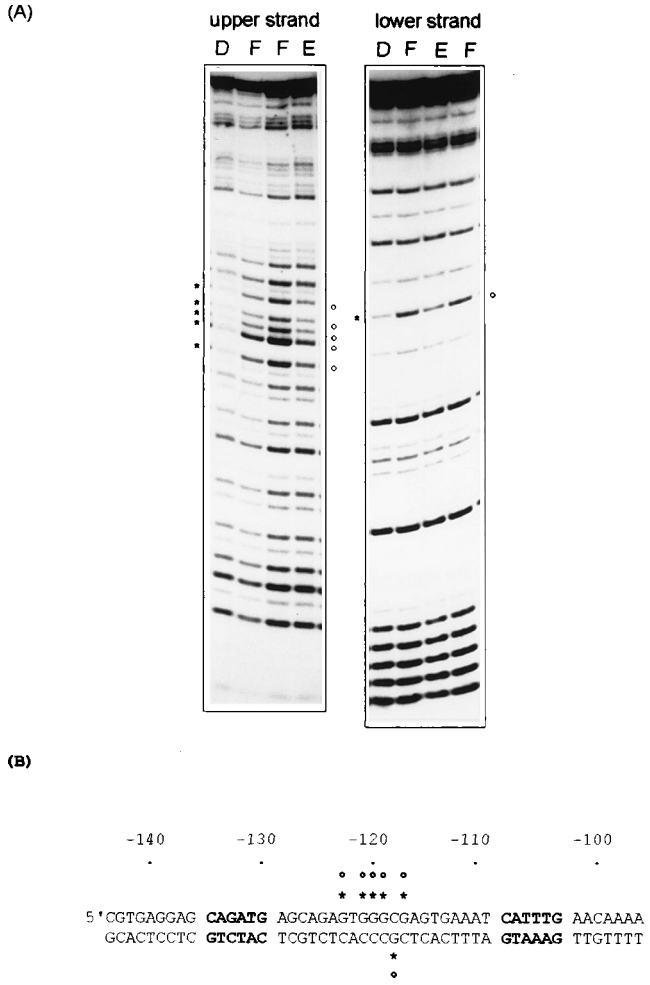

FIG. 11. Methylation interference analyses of the D and E complexes. The E-108/IS/E-130 oligonucleotide upper and lower strands were individually end labeled with ^{32}P and incubated with 60 μg of InR1G9 nuclear extracts as described in Materials and Methods. (A) Methylation interference patterns of D and E. Lanes F represent the cleavage patterns of free DNA, and lanes D and E represent the cleavage patterns of the D and E complexes, respectively. Modified methylated nucleotides are indicated along the ladder by stars (D) or circles (E). (B) Schematic representation of the methylation interference profiles of complexes D and E. Stars and circles indicate residues in D and E complexes, respectively, at which methylation specifically interfered with protein binding.

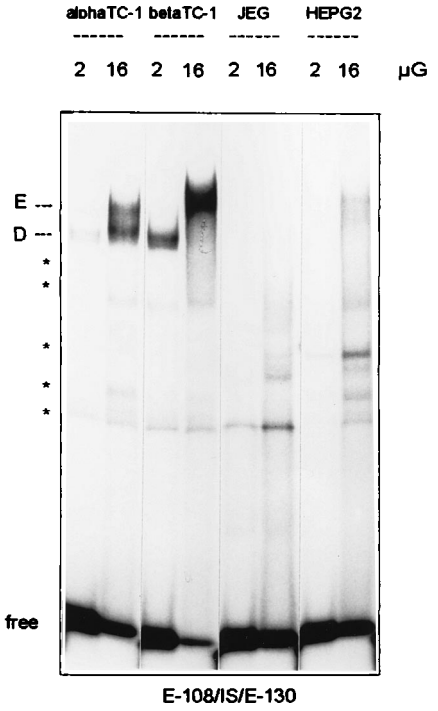

FIG. 12. Cell type distribution of the D and E complexes. Equal amounts of nuclear extracts (2 and 16 μg) from islet (alphaTC1 and betaTC1) and nonislet (JEG-3 and HepG2) cells were incubated with ^{32}P-labeled E-108/IS/E-130 oligonucleotides. EMSAs were performed as described for Fig. 3. D and E indicate the respective specific protein complexes, stars indicate nonspecific complexes, and free indicates free DNA.

proteins from InR1G9 cells and thus does not appear to have any function in the alpha-cell-specific expression of the glucagon gene (26). Actually, three complexes bind to a more extensive region within G1 between nt −75 and −96, which includes E1; the three complexes, however, do not exhibit any specificity for the E-box motif and are not affected by antibodies against the HLH protein E12/E47 (26). It is somewhat surprising that E1 cannot bind the A or IEF1 complexes, since it shares an identical core motif with E3 and the E boxes of the rat insulin I and II genes (Table 2). It has been suggested, however, that bases surrounding the CANNTG motif are important for determining the specificity of binding to different HLH proteins and that T residues flanking the motif may decrease or prevent their binding (13). PCR amplification of sequences bound by E47 failed to identify sequences with a T residue in that position (4); similarly, binding of the Myc protein is also inhibited by flanking T residues (14). The detrimental effect of flanking T residues on binding cannot be generalized to include all HLH proteins, however, inasmuch as binding of MyoD and Max does not appear to be affected (4, 14). The E1 motif is flanked on both the 5′ and 3′ sides by T residues; this characteristic may well prevent the binding of E12/E47-containing complexes. In that regard, it is interesting that the E3 motif has one flanking T residue and exhibits a lower affinity for the A/IEF1 complexes than does the NIR box of the rat insulin I gene, which has no immediately surrounding T residues.

In conclusion, the promoter of the glucagon gene can be divided into two control elements, each responsible for a specific level of expression. This double control may be necessary during development. When the first islet progenitor cells appear, they are able to coexpress multiple hormone genes (1, 16, 45); at that time, common *cis*-acting and *trans*-acting factors may be involved in the activation of these genes. The second level of specificity would intervene later at the stages of cell differentiation. Characterization of the complexes binding to G1 and G4 and studies of their temporal expression relative to that of the pancreatic hormone genes during embryogenesis should shed further light on the process of islet cell differentiation.

ACKNOWLEDGMENTS

We thank Isabel Pacheco for expert technical help and Florence Kaempfen for typing the manuscript. We also thank B. Mach and his group for helpful advice.

This work was supported by the Swiss National Science Foundation, the Helmut Horten and Wolfermann-Nägeli Foundations, the Institute for Human Genetics and Biochemistry, and the Sandoz and Ciba-Geigy Foundations and through an EMBO fellowship to M.C.-B.

REFERENCES

1. **Alpert, S., D. Hanahan, and G. Teitelman.** 1988. Hybrid insulin genes reveal a developmental lineage for pancreatic endocrine cells and imply a relationship with neurons. Cell **53:**295–308.
2. **Aronheim, A., H. Ohlsson, C. W. Park, T. Edlund, and M. D. Walker.** 1991. Distribution and characterization of helix-loop-helix enhancer-binding proteins from pancreatic beta cells and lymphocytes. Nucleic Acids Res. **19:** 3893–3899.
3. **Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub.** 1990. The protein Id: a negative regulator of helix-loop-helix proteins. Cell **61:**49–59.
4. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250:**1104–1110.
5. **Caudy, M., H. Vässin, M. Brand, R. Tuma, L. Y. Yan, and Y. N. Yan.** 1988. Daughterless, a drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the achaete-scute complex. Cell **55:**1061–1067.
6. **Decaux, J. F., B. Antoine, and A. Kahn.** 1989. Regulation of the expression of the L-type pyruvate kinase gene in adult rat hepatocytes in primary cultures. J. Biol. Chem. **264:**11584–11590.
7. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11:**1475–1489.
8. **Dorn, A., J. Bollekens, A. Staub, C. Benoist, and D. Mathis.** 1987. A multiplicity of CCAAT box-binding proteins. Cell **50:**863–872.
9. **Drucker, D. J., and P. L. Brubaker.** 1989. The rat glucagon gene is regulated by a CAMP-dependent pathway in rat intestine. Proc. Natl. Acad. Sci. USA **86:**3953–3957.
10. **Drucker, D. J., J. Philippe, L. Jepeal, and J. F. Habener.** 1987. Glucagon gene 5′-flanking sequences promote islet cell-specific glucagon gene transcription. J. Biol. Chem. **262:**15659–15665.
11. **Efrat, S., S. Linde, H. Kofod, D. Spector, M. Delannoy, S. Grant, D. Hanahan, and S. Baekkeskov.** 1988. Beta cell lines derived from transgenic mice expressing a hybrid insulin gene-oncogene. Proc. Natl. Acad. Sci. USA **85:** 9037–9041.
12. **Efrat, S., G. Teitelman, M. Anwar, D. Ruggiero, and D. Hanahan.** 1988. Glucagon gene regulatory region directs oncoprotein expression to neurons and pancreatic alpha cells. Neuron **1:**605–613.
13. **Fisher, F., and C. R. Goding.** 1992. Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. EMBO J. **11:**4103–4109.
14. **Halazonetis, T. D., and A. N. Kandil.** 1991. Determination of the C-MYC DNA-binding site. Proc. Natl. Acad. Sci. USA **88:**6162–6166.
15. **Henthorn, P., P. Zervos, M. Raducha, H. Harris, and T. Kadesch.** 1988. Expression of a human placental alkaline phosphatase gene in transfected cells: use as a reporter for studies of gene expression. Proc. Natl. Acad. Sci. USA **85:**6342–6346.
16. **Herrera, P. L., J. Huarte, F. Sanvito, P. Meda, L. Orci, and J. D. Vassali.** 1991. Embryogenesis of the murine endocrine pancreas: early expression of the pancreatic peptide gene. Development **113:**1257–1265.
17. **Hwung, J. P., J. Z. Gu, and M. J. Tsai.** 1990. Cooperativity of sequence elements mediates tissue specificity of the rat insulin II gene. Mol. Cell. Biol. **10:**1784–1788.
18. **Kadesch, T.** 1992. Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription. Immunol. Today **13:**31–37.
19. **Kadonaga, J., K. Jones, and R. Tjian.** 1986. Promoter-specific activation of the RNA polymerase II transcription by SP1. Trends Biochem. Sci. **11:**20–23.
20. **Karlsson, O., M. D. Walker, W. J. Rutter, and T. Edlung.** 1989. Individual protein binding domains of the insulin gene enhancer positively activate beta-cell transcription. Mol. Cell. Biol. **9:**823–827.
21. **Knepel, W., J. Chafitz, and J. F. Habener.** 1990. Transcriptional activation of the rat glucagon gene by the cAMP-responsive element in pancreatic islet cells. Mol. Cell. Biol. **10:**6799–6804.
22. **Kobr, M., W. Reith, C. Herrero-Sanchez, and B. Mach.** 1990. Two DNA-binding proteins discriminate between the promoters of different members of the major histocompatibility complex class II multigene family. Mol. Cell. Biol. **10:**965–971.
23. **Kruse, F., S. D. Rose, G. H. Swif, R. E. Hammer, and R. J. McDonald.** 1993. An endocrine-specific element is an integral component of an exocrine-specific pancreatic enhancer. Genes Dev. **7:**774–786.
24. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
25. **Merchant, J. L., B. Demediuk, and S. J. Brand.** 1991. A CG-rich element confers epidermal growth factor responsiveness to transcription from the gastrin promoter. Mol. Cell. Biol. **11:**2686–2696.
26. **Morel, C., M. Cordier-Bussat, and J. Philippe.** 1995. The upstream promoter element of the glucagon gene, G1, confers pancreatic alpha cell-specific expression. J. Biol. Chem. **270:**3046–3055.
27. **Murre, C., P. Schonleber McCow, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56:**777–783.
28. **Murre, C., P. Schonleber McCow, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to common DNA sequences. Cell **58:**537–544.
29. **Nelson, C., L. P. Shen, A. Meister, E. Fodor, and W. J. Rutter.** 1990. Pan: a transcriptional regulator that binds chymotrypsin, insulin and AP-4 enhancer motifs. Genes Dev. **4:**1035–1043.
30. **Ohlsson, H., O. Karlsson, and T. Edlund.** 1988. A beta-cell-specific protein binds to the two major regulatory sequences of the insulin gene enhancer. Proc. Natl. Acad. Sci. USA **85:**4228–4231.
31. **Ohlsson, H., S. Thor, and T. Edlund.** 1991. Novel insulin promoter- and enhancer-binding proteins that discriminate between alpha and beta cells. Mol. Endocrinol. **5:**897–904.
32. **Park, C. W., and M. D. Walker.** 1992. Subunit structure of cell-specific E box-binding proteins analyzed by quantitation of electrophoretic mobility shift. J. Biol. Chem. **267:**15642–15649.
33. **Philippe, J.** 1991. Structure and pancreatic expression of the insulin and glucagon genes. Endocrinol. Rev. **12:**252–271.
34. **Philippe, J., W. L. Chick, and J. F. Habener.** 1987. Multipotential phenotypic expression of genes encoding peptide hormones in clonal rat insulinoma cell lines. J. Clin. Invest. **79:**351–358.
35. **Philippe, J., D. J. Drucker, W. Knepel, L. Jepeal, Z. Misulovin, and J. F. Habener.** 1988. Alpha cell specific expression of the glucagon gene is conferred to the glucagon promoter element by the interactions of the DNA-binding proteins. Mol. Cell. Biol. **8:**4877–4888.
36. **Philippe, J., C. Morel, and M. Cordier-Bussat.** 1995. Islet-specific proteins interact with the insulin-response element of the glucagon gene. J. Biol. Chem. **270:**3039–3045.
37. **Philippe, J., A. C. Powers, S. Mojsov, D. Drucker, R. Comi, and J. F. Habener.** 1988. Expression of peptide hormone genes in human islet cell tumors. Diabetes **37:**1647–1651.
38. **Philippe, J., and S. Rochat.** 1991. Strict distance requirements for transcriptional activation by two regulatory elements of the glucagon gene. DNA Cell Biol. **10:**119–124.
39. **Powers, A. C., S. Efrat, S. Mojsov, D. Spector, J. F. Habener, and D. Hanahan.** 1990. Proglucagon processing similar to normal islets in pancreatic alpha-like cell lines derived from transgenic mouse tumors. Diabetes **39:**406–414.
40. **Prost, E., and D. D. Moore.** 1986. CAT expression vectors for analyses of eukaryotic promoters and enhancers. Gene **45:**107–111.
41. **Ruvkun, G., and M. Finney.** 1991. Regulation of transcription and cell identity by POU domain proteins. Cell **64:**475–478.
42. **Schaffner, W.** 1989. How do different transcription factors binding the same DNA sequence sort out their jobs? Trends Genet. **5:**37–39.
43. **Schreiber, E., P. Mathias, M. M. Muller, and W. Schaffner.** 1988. Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay. EMBO J. **7:**4221–4229.
44. **Takaki, R., J. Ono, M. Nakmura, M. Yokogama, S. Kumae, T. Hiraoka, K. Yamaguchi, K. Hamaguchi, and S. Uchida.** 1986. Isolation of glucagon-secreting cell lines by cloning insulinoma cells. In Vitro Cell. Dev. Biol. **22:**120–126.
45. **Teitelman, G., S. Alpert, J. M. Polak, A. Martinez, and D. Hanahan.** 1993. Precursor cells of mouse endocrine pancreas coexpress insulin, glucagon, and the neuronal proteins tyrosine hydroxylase and neuropeptide Y, but not pancreatic polypeptide. Development **118:**1031–1039.
46. **Therrien, M., and J. Drouin.** 1993. Cell-specific helix-loop-helix factor required for pituitary expression of the proopiomelanocortin gene. Mol. Cell. Biol. **13:**2342–2353.
47. **Voss, J. W., and M. G. Rosenfeld.** 1992. Anterior pituitary development: short tales from dwarf mice. Cell **70:**527–530.
48. **Walker, M. D., C. W. Park, A. Rosen, and A. Aronheim.** 1990. A cDNA from a mouse pancreatic beta cell encoding a putative transcription factor of the insulin gene. Nucleic Acids Res. **18:**1159–1166.
49. **Wang, T. C., and S. J. Brand.** 1990. Islet cell-specific regulatory domain in the gastrin promoter contains adjacent positive and negative DNA elements. J. Biol. Chem. **265:**8908–8914.
50. **Wheeler, M. B., J. Nishitani, A. M. J. Buchan, A. S. Kopin, W. Y. Chey, T. M. Chang, and A. B. Leiter.** 1992. Identification of a transcriptional enhancer important for enteroendocrine and pancreatic islet cell-specific expression of the secretin gene. Mol. Cell. Biol. **12:**3531–3539.
51. **Wright, W. E.** 1992. Muscle basic helix-loop-helix proteins and the regulation of myogenesis. Curr. Opin. Genet. Dev. **2:**243–248.