MOLECULAR AND CELLULAR BIOLOGY, Jan. 1996, p. 19–28 Vol. 16, No. 1
0270-7306/96/$04.00+0

Activation of Proglucagon Gene Transcription through a Novel Promoter Element by the caudal-Related Homeodomain Protein cdx-2/3

TIANRU JIN[1] AND DANIEL J. DRUCKER[2,3]*

Departments of Medicine,[1] Clinical Biochemistry,[2] and Molecular Genetics,[3] Banting and Best Diabetes Center, University of Toronto, Toronto, Ontario, Canada M5G 2C4

Received 18 April 1995/Returned for modification 9 June 1995/Accepted 17 October 1995

The proglucagon gene is expressed in a highly restricted tissue-specific manner in the A cells of the pancreatic islet and the L cells of the small and large intestines. The results of previous experiments indicate that cell-specific expression of the proglucagon gene is mediated by proteins that interact with the proximal G1 promoter element. We show here that the G1 element contains several AT-rich subdomains that bind proteins present in islet and enteroendocrine cell extracts. Electrophoretic mobility shift assay experiments using specific antisera identified the homeobox protein cdx-2/3 (which designates the same homeobox protein called cdx-2 for mice and cdx-3 for hamsters) as a major component of the G1-Gc2 complex in islet and intestinal cells. Mutations of the Gc element that decreased cdx-2/3 binding also resulted in decreased proglucagon promoter activity in islet and intestinal cell lines. The finding that cdx-2/3 mediates activation of the proglucagon promoter in both islet and enteroendocrine cells is consistent with the common endodermal lineage of these tissues and provides new insight into the coordinate regulation of genes expressed in both pancreatic and intestinal endocrine cell types.

The proglucagon gene is expressed in the A cells of pancreatic islets, selected neurons of the hypothalamus and brain stem, and enteroendocrine cells of the small and large intestines (1, 5, 6). Proglucagon gene transcription results in the production of a single proglucagon mRNA transcript that is identical in all three tissues (23). Posttranslational processing of proglucagon is highly tissue specific, resulting in the liberation of a unique profile of proglucagon-derived peptides in the pancreas, intestine, and brain (23, 25). Glucagon appears to be the most important biologically active proglucagon-derived peptide liberated in the pancreatic islets, whereas processing of proglucagon in enteroendocrine cells of the small and large intestines results in the liberation of several bioactive peptides, including oxyntomodulin, GLP-1, and glicentin, as well as GLP-2 and two intervening peptides (23).

Experiments using islet cell lines have led to the elucidation of both the signal transduction pathways and the *cis*-acting elements that mediate regulation of islet cell-specific proglucagon gene transcription (8, 27). Two enhancer-like elements (designated G3 and G2), a cyclic AMP response element and an islet cell-specific promoter element (G1), have been identified within the first 300 bp of the rat proglucagon gene 5′-flanking sequence (15, 16, 27). More detailed analyses of G3 and G2 subdomains have identified specific DNA sequences that mediate inhibition of proglucagon gene transcription by insulin and HNF-3β, respectively (29, 30). Whereas G2 and G3 appear to display enhancer-like properties, the cell-specific promoter activity of the proglucagon gene has been mapped to the more proximal G1 element. G1 sequences interact predominantly with proteins from islet nuclear extracts (24), suggesting that the highly restricted cellular specificity of the proglucagon promoter is mediated by the interaction of islet cell-specific proteins with subdomains of the proximal G1 promoter element (24).

Analysis of the transcriptional and DNA-binding properties of the proglucagon gene G1 region (which extends from bp −100 to −52) has localized the determinants of islet cell-specific proglucagon gene transcription (24) to a specific region containing two TAAT motifs characteristic of the core elements that bind homeobox transcription factors. These observations suggest that the islet cell-specific expression of the proglucagon gene may be mediated by one or more homeobox proteins expressed in the endocrine A cells of the pancreatic islets. Although the homeobox protein STF-1/IPF-1/IDX-1 activates the insulin promoter and is important for pancreas formation, the lack of *STF-1*/*IPF-1*/*IDX-1* expression in the majority of glucagon-producing islet cells (18, 20, 26) suggests that other, possibly novel, homeobox proteins regulate the transcriptional control of the proglucagon promoter through the G1 element.

Identification of homeobox transcription factors that bind to islet hormone gene promoters has been undertaken by using a number of different experimental approaches. The insulin gene far-linked AT-rich enhancer sequence was used as a probe in expression-cloning studies to isolate hamster far-linked AT-rich element-binding proteins, including the LIM-domain protein lmx-1 and the caudal-related protein cdx-3 (11). An alternative approach for the identification of islet homeobox sequences makes use of the high degree of conservation of amino acid sequences within helix 3 and helix 1 of the homeobox protein. This PCR-based strategy resulted in the isolation of several novel homeobox sequences from rat islet cDNA, including the *STF-1*/*IPF-1*/*IDX-1* gene, and sequences corresponding to members of the *caudal* gene family (20). A similar experimental approach was used to isolate homeobox DNA sequences from a hamster insulinoma cell cDNA library, including eight genes not previously known to be expressed in islet cells (33).

The restriction of proglucagon gene expression to islets and

* Corresponding author. Mailing address: The Toronto Hospital, 200 Elizabeth St., CCRW3-838, Toronto, Ontario, Canada M5G 2C4. Phone: (416) 340-4125. Fax: (416) 978-4108.

intestine suggested that transcriptional regulators expressed in both the A cells of the endocrine pancreas and the L cells of the gastrointestinal tract may be good candidates for *trans*-acting factors that control proglucagon gene transcription through the G1 promoter element. As members of the caudal-related homeobox gene family play a role in early *Drosophila* development, are expressed during gastrulation and in intestinal endoderm (12, 21, 22), and have been identified in both islets and intestine, we reasoned that these genes were good candidates for regulation of cell-specific proglucagon gene transcription through the AT-rich sequences in the proglucagon promoter. The hamster *cdx-3* cDNA (11) isolated from islet cells shares over 90% amino acid homology with the mouse *cdx-2* sequence cloned from the intestine (14, 36), suggesting that these two genes encode homologous proteins in different species. We now report that the homeobox protein cdx-2/3 (which designates the same homeobox protein called cdx-2 for mice and cdx-3 for hamsters) mediates the cell-specific expression of the proglucagon gene in islets and intestine by specifically activating the G1 element of the proglucagon promoter.

MATERIALS AND METHODS

Plasmids. The plasmid phcdx-3 (also called pBAT7.cdx-3), as well as cdx-3 antiserum, was kindly provided by M. S. German, San Francisco, Calif. The expression of hamster cdx-3 protein (313 amino acids) in this plasmid is under the control of the cytomegalovirus (CMV) promoter. An antisense control plasmid, cdx-3(AS), was generated by cloning the *cdx-3* sequence in the 3′ to 5′ orientation with the two *Not*I restriction sites in phcdx-3. Rat proglucagon gene 5′-flanking sequences from bp −2292 to +58 (17) were subcloned into the promoterless plasmid pBLuc immediately adjacent to the coding sequences of the firefly luciferase reporter gene. This plasmid was designated (−2292) GluLuc. A series of 5′-deleted proglucagon-luciferase fusion genes (from bp −2292 to −60) was generated with exonuclease III. The precise sequence at the 5′ end of each deletion was confirmed by DNA sequencing. CMV-Luc contains the CMV promoter adjacent to the luciferase coding sequence in pBLuc. SMS-Luc was generated by inserting an 800-bp *Xba*I fragment of the rat somatostatin (SMS) promoter (31) upstream of the luciferase coding sequence in pBLuc.

Cell culture, RNA analysis, and transfection. All cell lines were grown in Dulbecco's modified Eagle's medium (4.5 g of glucose per liter). Hamster BHK fibroblasts and InR1-G9 islet cells (9) were supplemented with 5% calf serum. Rat RIN1056A islet cells (28) were supplemented with 10% calf serum, and mouse intestine STC-1 and GLUTag cell lines (7, 32) were supplemented with 2.5% fetal bovine serum plus 10% horse serum or 10% fetal bovine serum, respectively. BHK cells and InR1-G9 cells were transfected by the calcium phosphate precipitation method with a glycerol shock 4 h after the transfection. STC-1 cells and GLUTag cells were transfected by electroporation. The cells were harvested for analysis 16 to 20 h after transfection. The luciferase activity was analyzed as described previously (4), and values were normalized to the protein concentration in each extract. The luciferase activity of the specific reporter plasmids was normalized relative to the background luciferase activity obtained following transfection of the promoterless luciferase plasmid pBLuc in the same experiment.

RNA was prepared by the acid ethanol precipitation method as previously described (2), and RNA size separation by gel electrophoresis, blotting, hybridization, and washing was carried out as previously described (6, 7).

Nuclear protein extraction and EMSA. Nuclear proteins from three proglucagon-producing cell lines, InR1-G9, STC-1, and GLUTag, and BHK fibroblasts were prepared as previously described (35). Synthetic oligonucleotides corresponding to specific G1 sequences were annealed, separated by electrophoresis with a 5% polyacrylamide gel, and ^{32}P labeled with the Klenow fragment of DNA polymerase. The electrophoretic mobility shift assay (EMSA) was performed by incubating 4×10^4 cpm of end-labeled DNA probe (0.02 to 0.03 ng of DNA) with nuclear protein (0.75 to 7.5 μg) in a binding buffer (10 mM Tris-HCl [pH. 8.0], 40 mM KCl, 6% glycerol, 1 mM dithiothreitol, 0.05% Nonidet P-40) for 30 min at 30°C. For supershift experiments, nuclear proteins were premixed with diluted preimmune or cdx-3-specific antiserum in the binding buffer at 0°C for 20 min before the addition of labeled DNA probe. The reaction mixture was then loaded onto a 5% nondenaturing polyacrylamide gel, and following electrophoresis, the gel was exposed to X-ray film for 16 to 24 h.

DNase I footprinting assay. Nuclear extract from the mouse large-bowel GLUTag cell line was incubated with ~5 ng (2×10^4 cpm) of 5′-end-labeled rat proglucagon gene (bp −155 to +58) fragment in a volume of 30 μl (in 1× EMSA binding buffer in the presence of 1 μg of poly(dI-dC). The incubation was carried out at 28°C for 30 min. The DNase I digestion (0.3 to 1.2 ng) was carried out by using the Sure Track Footprinting Kit (Pharmacia Biotech) as previously described (27).

Western blot (immunoblot) analysis. Twenty-five-microgram portions of nuclear extracts from various cell lines were size fractionated on a sodium dodecyl sulfate (SDS)–12% polyacrylamide gel and electrophoretically transferred onto a nitrocellulose filter. cdx-2/3 immunoreactive protein was detected with the ECL Western analysis system (Amersham Life Sciences) with the peroxidase-linked anti-rabbit immunoglobulin as the second antibody.

RESULTS

***cdx-2/3* expression in islet and intestinal endocrine cells.** Analysis of homeobox gene expression in islet and intestinal cells by reverse transcription-PCR revealed that the *cdx-2/3* gene was expressed in both tissues (20, 33). To determine if expression of *cdx-2/3* was associated with proglucagon gene expression in different cell types, we analyzed islet and intestinal proglucagon-producing cell lines for the presence of cdx-2/3 mRNA transcripts. Using a 550-bp fragment of the hamster *cdx-3* 3′ coding sequences as a probe in Northern (RNA) blot analysis (11), we detected cdx-2/3 mRNA transcripts of the predicted size in four proglucagon-producing cell lines, the mouse small-intestine STC-1 and large-intestine GLUTag cells and the hamster islet InR1-G9 and rat islet RIN1056A cells, and in one hamster insulinoma cell line, HIT cells (Fig. 1); cdx-2/3 mRNA transcripts in RIN1056A cells were more readily observed after a longer exposure of the blot shown in Fig. 1. No cdx-2/3 mRNA transcripts were observed with RNA isolated from NIH 3T3 or BHK fibroblasts or from In111 hamster insulinoma cells (Fig. 1). The absence of cdx-2/3 mRNA in various cell lines was not simply attributable to degradation of RNA, as indicated by the intact 28S and 18S ribosomal bands and the roughly comparable amounts of tubulin mRNA transcripts detected after rehybridization of the same blot with a tubulin cDNA probe (Fig. 1).

To ascertain whether the cdx-2/3 mRNA transcripts detected in the proglucagon-producing cell lines are translated to cdx-2/3 immunoreactive protein, we prepared extracts from these cell lines for Western analysis with cdx-2/3 antisera. A major cdx-2/3 immunoreactive band of ~35 Da was detected in InR1-G9, GLUTag, and STC-1 cells. In contrast, cdx-2/3 was not detected in wild-type BHK cells; however, following transfection of BHK cells with a *cdx-2/3* expression vector, a cdx-2/3 immunoreactive band that comigrated with the endogenous cdx-2/3 immunoreactive protein detected in the proglucagon-producing cell lines was observed. Taken together, these data indicate that cdx-2/3 mRNA transcripts and immunoreactive protein are expressed in proglucagon-producing islet and intestinal cell lines.

Hamster cdx-3 activates the proglucagon promoter in BHK fibroblasts. The detection of cdx-2/3 mRNA transcripts in all four proglucagon-producing cell lines suggested that this gene may be a candidate regulator of proglucagon gene transcription. To examine this hypothesis, we transfected BHK fibroblasts (which do not contain cdx-3 mRNA transcripts) with a series of proglucagon promoter-luciferase fusion genes (Fig. 2a). As expected, the level of transcriptional activity of the proglucagon-luciferase plasmids was low in BHK cells; cotransfection of the proglucagon promoter-luciferase plasmids and hamster *cdx-3* resulted in a 6- to 15-fold activation of the proglucagon promoter (Fig. 2a). The cdx-3-dependent activation of the proglucagon promoter was observed with serially deleted fusion genes containing the deletion of bp −2292 to −82 of rat proglucagon gene 5′-flanking sequences. In contrast, further deletion to bp −60 resulted in a loss of cdx-3 activation. No activation of luciferase activity by cdx-3 was seen with the promoterless luciferase plasmid pBLuc (data not

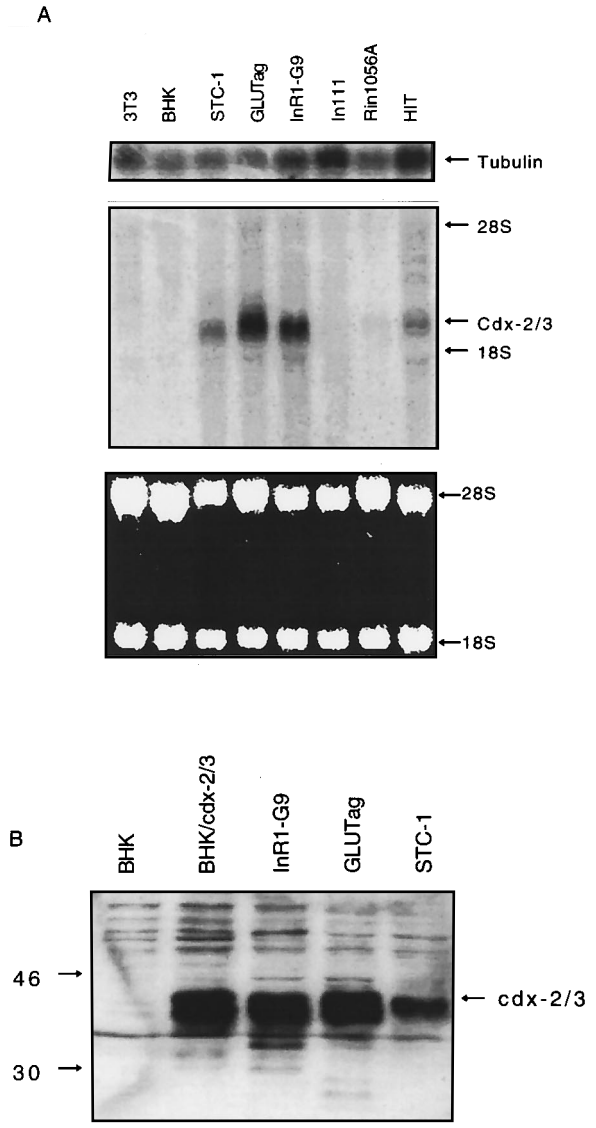


FIG. 1. Northern blot analysis of cdx-2/3 in islet and intestinal cell lines. Ten micrograms of RNA from Swiss NIH 3T3 and BHK fibroblasts, small-bowel STC-1 cells, large-bowel GLUTag cells, islet InR1-G9, In111, and RIN1056A cells, and insulinoma HIT cells were size fractionated on an agarose gel, transferred to a nylon membrane, and hybridized with a hamster *cdx-3* cDNA containing 500 bp of hamster *cdx-3* 3′ coding sequences. Integrity of the RNA and equality of loading were assessed by staining of the gel with ethidium bromide prior to transfer (bottom panel) and rehybridization of the blot with a tubulin cDNA probe (top panel). The migration positions of the 28S and 18S ribosomal RNAs are indicated by arrows. (B) Expression of cdx-2/3 protein in proglucagon-producing cells. Twenty-five-microgram portions of nuclear extract from three proglucagon-producing cell lines (InR1-G9, GLUTag, and STC-1), wild-type BHK fibroblasts, and BHK cells transfected with *cdx-2/3* were size fractionated on an SDS–12% polyacrylamide gel and transferred to a nylon membrane electrophoretically. cdx-2/3 immunoreactive protein was detected with a rabbit antibody directed against the hamster cdx-3 C-terminal domain (a gift from M. German) and visualized by using peroxidase-linked anti-rabbit immunoglobulin and the ECL Western analysis system (Amersham Life Sciences). The migration position of cdx-2/3 protein is slightly higher than the predicted molecular mass (33.5 kDa) based on the *cdx-2/3* open reading frame.

shown). Furthermore, cotransfection with either the CMV expression vector (pBAT7) alone or the CMV expression vector with *cdx-3* coding sequences in the antisense orientation, *cdx-3*(AS), did not result in activation of the proglucagon promoter (Fig. 2a). To assess the promoter specificity of the cdx-3 activation, we cotransfected *cdx-3* with luciferase plasmids containing (i) a viral promoter (CMV-Luc) or (ii) the promoter (SMS) of a gene that is also expressed in the islets and intestine and is known to be regulated by homeobox proteins (SMS-Luc). In contrast to the cdx-3-dependent activation of proglucagon promoter sequences, cdx-3 did not activate the CMV or somatostatin promoters (Fig. 2b and c).

To determine whether additional homeobox-containing transcription factors were also capable of activating the bp −82 proglucagon promoter, we transfected BHK cells with the (−82) Glu-Luc plasmid and the identical CMV expression vector pBAT7 containing the homeobox transcription factors pit-1 or isl-1. In contrast to the cdx-3-dependent activation, no activation of the proglucagon promoter was detected following cotransfection of the isl-1 or pit-1 plasmids (Fig. 2d).

cdx-3 activation of the proglucagon promoter correlates with binding to proximal G1 promoter sequences from bp −82 to −60. As shown in Fig. 2a, deletion of proglucagon gene sequences from bp −2292 to −82 had no effect on cdx-3 activation of the proglucagon promoter. However, further deletion to bp −60 completely abolished the activation by cdx-3, suggesting that the DNA sequences from bp −82 to −60 of the rat proglucagon promoter contain target sequences for cdx-3 binding and transactivation. Analysis of DNA-protein interactions by DNAse I footprinting with the −155 to +58 proglucagon gene fragment and increasing amounts of GLUTag nuclear extract demonstrated protection of the region from −80 to −65 on the coding strand and from −66 to −76 on the noncoding strand, consistent with previous observations (24, 27). This region of the proximal proglucagon gene promoter (Fig. 3a) revealed a number of AT-rich sequences that may serve as potential targets for cdx-3 binding. Furthermore, these sequences reside within the G1 domain of the proglucagon promoter, a region shown to be critical for cell-specific control of proglucagon gene transcription (24, 27).

To more precisely identify the sequences within G1 important for cdx-3 activation, we prepared four synthetic oligonucleotides that spanned this region (Ga, Gb, Gc, and Gd; Fig. 3c and d) and used these probes in EMSA experiments. Several specific DNA-protein complexes were generated with all four probes and InR1-G9 nuclear extracts (Fig. 3c and d). The formation of complexes Gc1 through Gc4 was diminished by competition with excess unlabeled competitor (see Fig. 4c) (data not shown). To determine whether cdx-3 was one of the proteins present in the Gc-Gd complexes, we carried out EMSA experiments in the presence of either preimmune or cdx-3-specific antisera. Antisera directed against the carboxy-terminal sequences of cdx-3 had no effect on the formation of the major binding complexes observed with the Ga or Gb probe (Fig. 3c, lanes 1 to 6). In contrast, of the four major complexes observed with the Gc probe, the Gc2 band was supershifted by the addition of anti-cdx-3 antiserum (Fig. 3c, lane 10). No change in the mobility of the Gc2 complex was observed with preimmune serum in the same experiment (Fig. 3c, lane 9). Similarly, the Gd probe, which is highly similar to Gc but contains a 4-bp extension at the 5′ end, generates a series of complexes with the InR1-G9 extract (Gd1 through Gd4) that comigrate with the complexes observed with the Gc element (Fig. 3d). Furthermore, cdx-3 antisera diminish the formation of the Gd2 complex and produce a supershifted complex similar to that obtained with the Gc probe (Fig. 3c). Similarly, BHK cells transfected with *cdx-3* also generate a Gd2 complex in EMSA experiments that is diminished and supershifted in the presence of cdx-3 antisera. These observations clearly demonstrate that subdomains of the G1 promoter sequence (as represented by oligonucleotides Ga, Gb, Gc, and

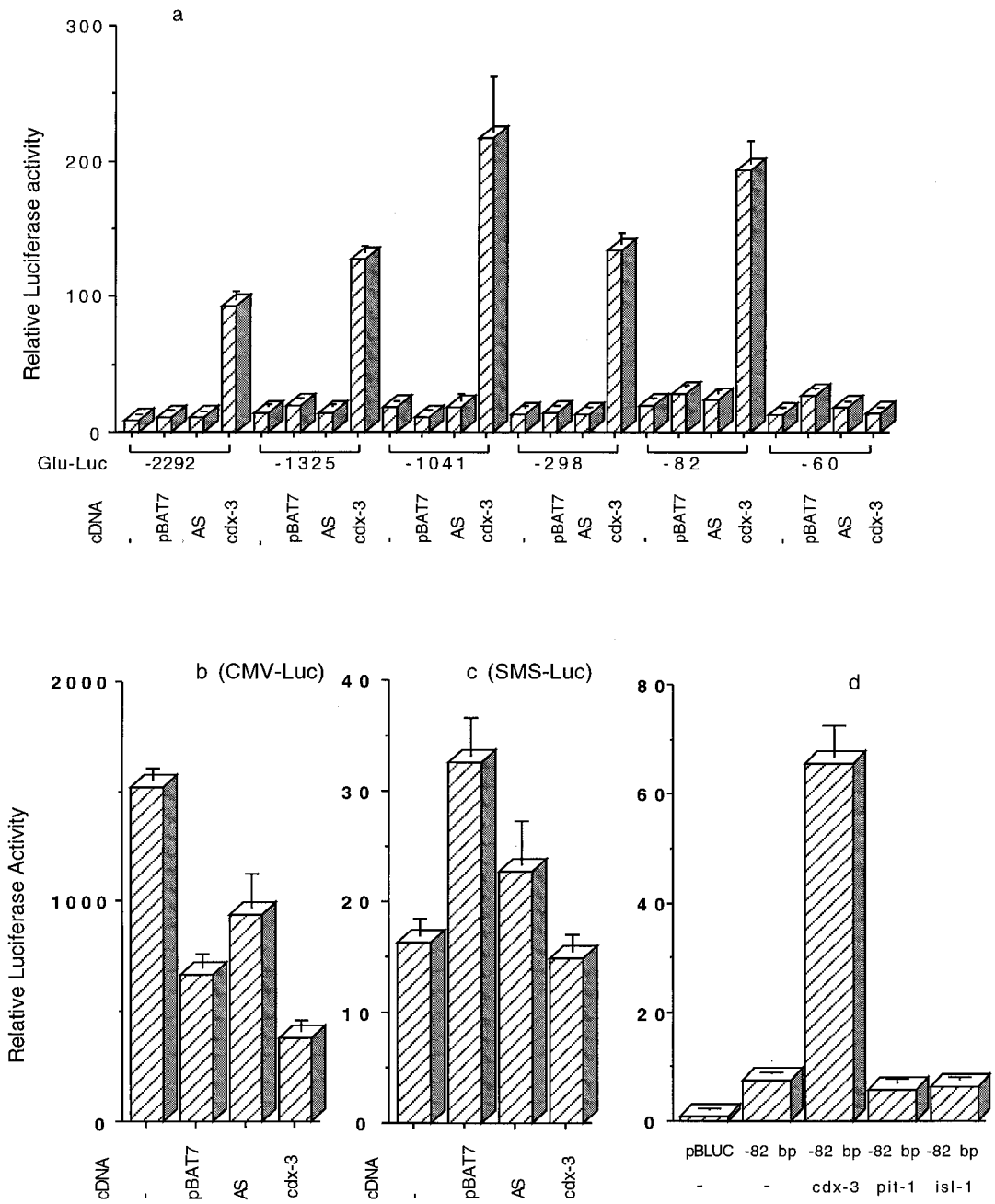


FIG. 2. cdx-3 activates the proglucagon promoter in BHK fibroblasts. (a) A series of proglucagon-luciferase fusion genes was transfected into BHK cells alone (−) or with the expression vector pBAT7, the *cdx-3* cDNA cloned into pBAT7 in the antisense orientation (AS), or the *cdx-3* sequences cloned in pBAT7 in the sense orientation (cdx-3). Ten micrograms of DNA of the reporter plasmid was transfected into BHK cells by the calcium phosphate precipitation method with 10 μg of plasmid DNA or 10 μg of pBAT7, *cdx-3*(AS), or *cdx-3*. The data are expressed as mean relative luciferase activity ± standard error of the mean normalized to the activity obtained following transfection of the promoterless luciferase plasmid in the same experiment. (b) BHK cells were transfected with CMV-Luc alone, or CMV-Luc was cotransfected with the CMV expression vector pBAT7 or with *cdx-3* subcloned into pBAT7 in the antisense (AS) or sense (cdx-3) orientation. (c) BHK cells were transfected with SMS-Luc alone, or SMS-Luc was cotransfected with the CMV expression vector pBAT7 or with *cdx-3* subcloned into pBAT7 in the antisense (AS) or sense (cdx-3) orientation. (d) BHK cells were transfected with pBLuc alone, or 10 μg of the (−82) Glu-Luc reporter plasmid was cotransfected with 10 μg of the pBAT7 plasmid containing cDNAs for the homeodomain proteins cdx-3, pit-1, and isl-1.

Gd) form complexes with comparable mobility that may contain different proteins. Furthermore, the data demonstrate that cdx-3, present in InR1-G9 islet cell extracts, binds to sequences within the proglucagon promoter Gc-Gd element to generate a Gc2-Gd2 complex.

As proglucagon gene transcription is highly tissue specific and restricted principally to islet and intestinal endocrine cells, we hypothesized that cdx-2/3 may contribute to the regulation of proglucagon promoter G1 activity in both pancreatic and intestinal cell lineages. Accordingly, we prepared nuclear extracts from two enteroendocrine cell lines, small-bowel STC-1 cells (32) and large-bowel GLUTag cells (7), and compared the Gc complexes formed with enteroendocrine and InR1-G9 islet extracts (Fig. 4). The four complexes detected with the Gc

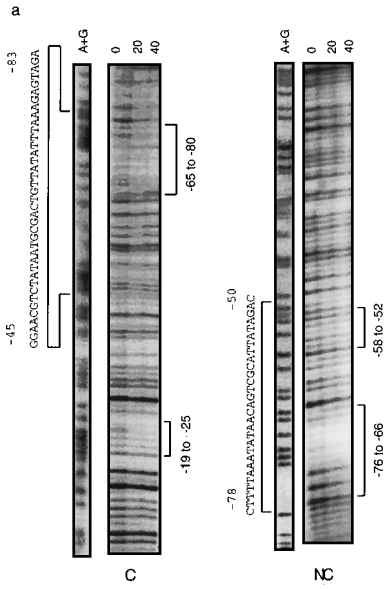


FIG. 3. cdx-3 binds to the Gc subdomain (bp −74 to −50) of the G1 element in the rat proglucagon promoter. (a) DNAse I footprinting with GLUTag nuclear extracts demonstrates protection of the G1 region. The rat proglucagon gene fragment from bp −155 to +58 was individually end labeled on both upper (C) and lower (NC) strands, and after purification, each labeled single-stranded fragment was incubated with DNase I and either 0, 20, or 40 μg of GLUTag nuclear extracts. G+A sequencing reactions and DNase I digestion were carried out with a Sure Track Footprinting Kit (Pharmacia Biotech). (b) The proximal rat proglucagon promoter sequences and the locations of the four G1 AT-rich subdomains are shown (27). (c) EMSA using islet InR1-G9 nuclear extract to identify G1 binding proteins. Nuclear extracts (7.5 μg) were incubated with the G1 probes shown in panel a either alone (−) or with 5 μl of preimmune antisera (pi) or 5 μl of 1/10 diluted cdx-3 antibody (cdx-3), following which DNA-protein complexes were resolved by polyacrylamide gel electrophoresis and autoradiography. F, migration position of unbound probe. Four major DNA-protein complexes observed were designated Gc1 to Gc4. Gc2/Ab, the supershifted complex observed in the presence of cdx-3 antisera. NS, nonspecific complex not diminished by the presence of excess specific competitor DNA. (d) EMSA identifies DNA-protein interactions with the probe Gd (bp −78 to −51). The Gd oligonucleotide has a 4-bp 5′ extension, compared with the 5′ end of Gc. Nuclear extract (2.5 μg) from *cdx-2/3*-transfected BHK cells or 7.5 μg of nuclear extract from InR1-G9 cells was incubated with the Gd probe alone (−) or with 5 μl of 1/10 diluted antiserum (preimmune [pi] or cdx-3 antiserum) as indicated. Four major DNA-protein complexes were designated Gd1 to Gd4.

b

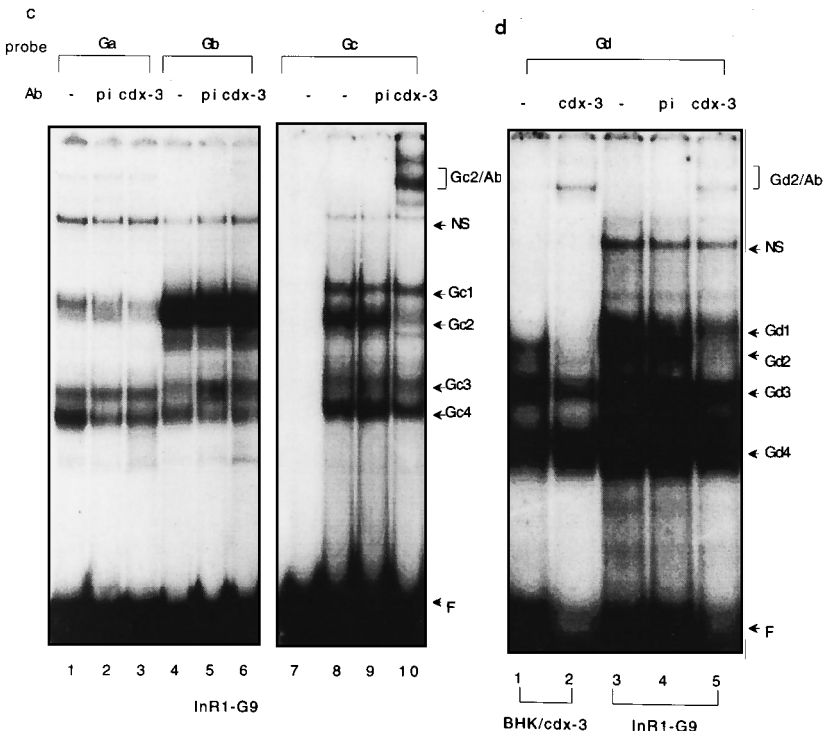
```
-94                                                    -37
CCATTATTTACAGATGAGAAATTTATATTGTCAGCGTAATATCTGCAAGGCTAAACAG
GGTAATAAATGTCTACTCTTTAAATATAACAGTCGCATTATAGACGTTCCGATTTGTC

-94        Gb          -70
CCATTATTTACAGATGAGAAATTTA
GGTAATAAATGTCTACTCTTTAAAT

                -74        Gc          -50
                ATTTATATTGTCAGCGTAATATCTG
                TAAATATAACAGTCGCATTATAGAC

                                  -61        Ga          -37
                                  GCGTAATATCTGCAAGGCTAAACAG
                                  CGCATTATAGACGTTCCGATTTGTC

          -78        Gd          -51
          AGAAATTTATATTGTCAGCGTAATATCT
          TCTTTAAATATAACAGTCGCATTATAGA
```



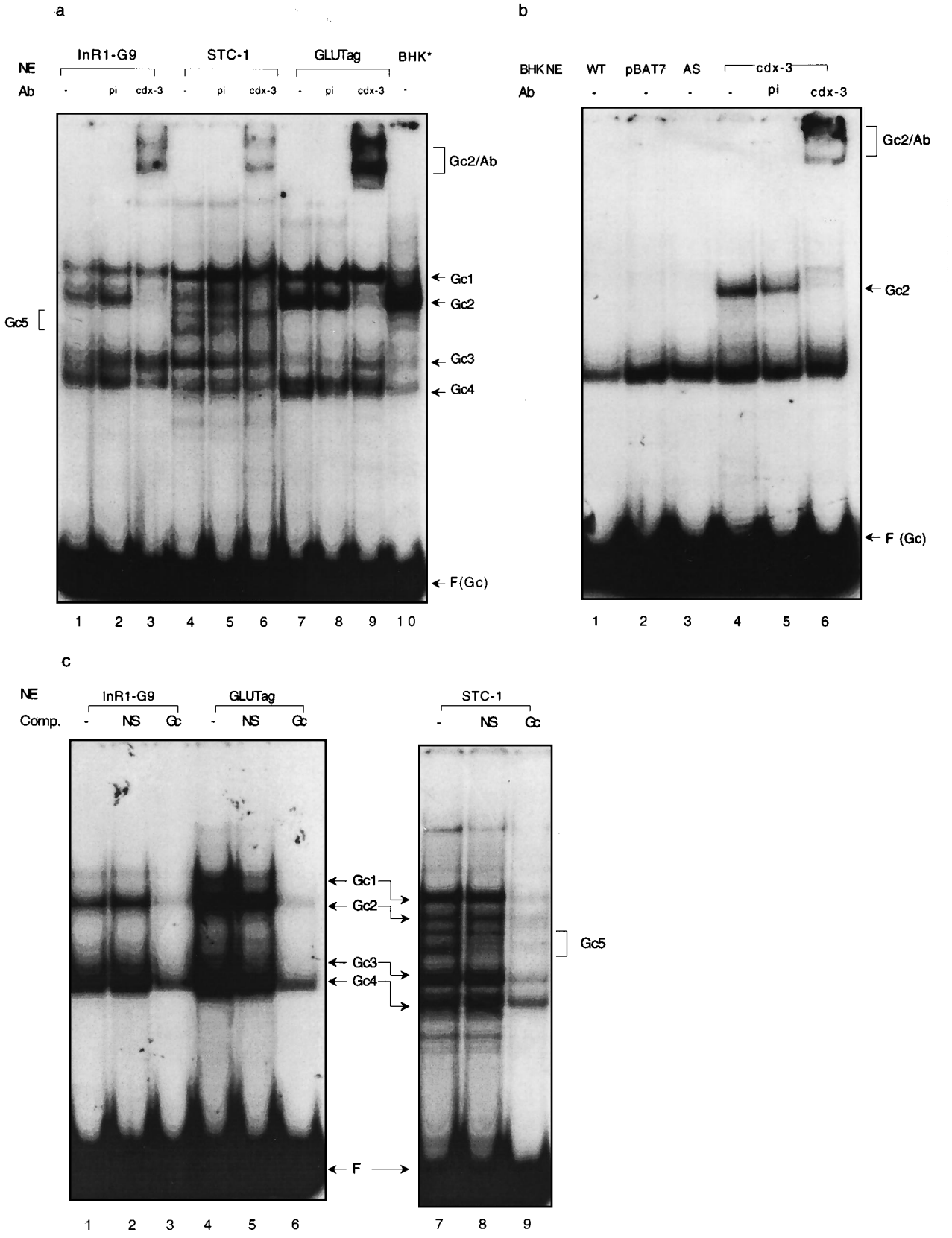

a
NE
InR1-G9
STC-1
GLUTag
BHK*
Ab
-
pi
cdx-3
Gc2/Ab
Gc1
Gc2
Gc5
Gc3
Gc4
F(Gc)
1 2 3 4 5 6 7 8 9 10
b
BHK NE
WT
pBAT7
AS
cdx-3
Ab
Gc2/Ab
Gc2
F (Gc)
1 2 3 4 5 6
c
NE
InR1-G9
GLUTag
STC-1
Comp.
NS
Gc
Gc1
Gc2
Gc3
Gc4
F
Gc5
1 2 3 4 5 6 7 8 9

FIG. 4. Detection of cdx-3 binding to the Gc element in islet and enteroendocrine cells. (a) EMSA shows the specific interaction of the Gc element and nuclear extract (7.5 μg) from InR1-G9 islet and STC-1 and GLUTag intestinal endocrine cells. Five major DNA-protein complexes observed were designated Gc1 to Gc5. Nuclear extract (NE) was incubated with Gc alone (−) or with 5 μl of 1/10 diluted preimmune (pi) or cdx-3 antisera before the addition of labeled Gc element. Ab, antibody; BHK*, BHK cells transfected with a *cdx-3* expression vector; Gc2/Ab, the supershifted complex observed in the presence of cdx-3 antisera; F(Gc), free probe. (b) EMSA shows the interaction of the Gc element with nuclear extracts from wild-type BHK cells (WT) and BHK cells transfected with the CMV expression vector alone (pBAT7), the antisense *cdx-3* sequences (AS), or *cdx-3*. NE (0.75 μg) was incubated with or without (−) 5 μl of 1/10 diluted cdx-3 antibody (cdx-3) or with preimmune sera (pi) before the addition of labeled Gc element. Gc2 denotes the appearance of a complex with migration characteristics comparable to those of the Gc2 complex detected with islet and intestinal extracts. Gc2/Ab, the supershifted complex observed in the presence of cdx-3 antisera. F(Gc), migration position of unbound probe. (c) Competition for protein binding to the Gc element using islet and enteroendocrine cell NEs. The EMSA binding reactions were carried out in the presence of a 200-fold excess of specific competitor Gc sequences (Gc) or a 1,000-fold excess of nonspecific competitor sequences (NS) (a 31-bp sequence from position −1680 to −1649 of the rat proglucagon gene 5′-flanking sequences). The positions of the Gc complexes and free probe (F) are indicated. Comp, competitor sequences.

probe when InR1-G9 extracts were used (Gc1 through Gc4) were also observed, albeit with some differences in relative abundance, with extracts from STC-1 and GLUTag cells. The formation of the Gc1 through Gc4 complexes was clearly diminished in the presence of excess specific Gc unlabeled competitor DNA but not by excess nonspecific competitor DNA (Fig. 4c). A faint complex, designated Gc5, was formed with STC-1 nuclear extracts but not with protein from InR1-G9 or GLUTag cells (Fig. 4a). To ascertain whether cdx-2/3 also contributed to the formation of the Gc2 complex in intestinal cells, we carried out EMSA experiments with preimmune and cdx-3-specific antisera (Fig 4a). The cdx-3 antisera produced a similar supershift of the Gc2 complex formed with nuclear extracts from all three glucagon-producing cell lines (Fig. 4a), consistent with the presence of cdx-2/3 in the Gc2 complex generated with both islet and intestinal extracts.

Formation of the Gc2 complex in BHK cells transfected with *cdx-3*. The observation that cdx-3 activates proglucagon promoter activity in fibroblasts, taken together with the detection of a specific cdx-2/3–Gc2 complex in glucagon-producing cells, suggested that formation of the cdx-2/3–Gc2 complex may be an important determinant for activation of the proglucagon promoter. To ascertain whether the low basal level of proglucagon gene transcription in fibroblasts was associated with a lack of cdx-3 activity, we prepared nuclear extracts (for EMSA experiments) from wild-type BHK cells (Fig. 4b, lane 1) and cells transfected with the expression vector (without cdx-3) alone (lane 2) or the expression vector containing *cdx-3* coding sequences in an antisense orientation (lane 3) and from BHK cells transfected with *cdx-3* (lanes 4 to 6). A specific Gc2 complex that comigrates with the Gc2 complex detected with islet and intestinal extracts (Fig. 4a, lane 10) was observed only in the BHK cells transfected with *cdx-3* (lanes 4 to 6). Furthermore, this complex was supershifted by cdx-3 antisera, but not by preimmune sera, providing further evidence that cdx-3 introduced into BHK cells not only transactivates the proglucagon promoter but is sufficient for the formation of the Gc2 complex in vitro.

Identification of the cdx-2/3 binding site in the proglucagon G1 region. To more precisely localize the nucleotide sequences within the Gc subdomain that were critical for cdx-2/3 binding to the proglucagon promoter, we synthesized a series of block mutations in the Gc sequence (Fig. 5a) and tested the binding properties of these mutations in EMSA experiments (Fig. 5b). Both the M1 and M2 mutants (which contain substitutions in the 5′ AT-rich domains) displayed markedly diminished Gc2 complex formation (Fig. 5b). The M3 sequence, which contains a block mutation in the middle of the Gc subdomain, displayed intact Gc2 complex formation. Interestingly, although M1, M2, and M3 displayed differences in their Gc2 binding properties, formation of the Gc4 complex was enhanced with all three mutants, despite their differential effects on formation of the Gc2 complex. Furthermore, the M4 mutant displayed a modest reduction in Gc2 formation, yet this mutation essentially eliminated formation of the Gc4 complex, whereas mutant M5 did not result in impaired formation of Gc2; however, formation of Gc3 was clearly attenuated. Similar results were obtained with the M1 through M5 mutants and GLUTag intestinal extracts (data not shown). Furthermore, the relative diminution of Gc2 complex formation observed with the various M1 through M5 mutants correlated with a reduction in the intensity of the Gc2 supershift detected with cdx-3 antisera in EMSA experiments with the M1 through M5 probes and both InR1-G9 and GLUTag extracts (data not shown). A Gc probe (M6) with more limited replacement nucleotide substitutions in the 5′ and 3′ AT-rich Gc subdomains (Fig. 5a) demonstrated intact formation of Gc3 and G4 with InR1-G9, GLUTag, and STC-1 extracts, but formation of Gc2 was clearly diminished with the M6 probe (Fig. 5c and data not shown). Taken together, these data provide evidence for differential nucleotide binding specificity of the various InR1-G9 and GLUTag proteins that interact with the proglucagon gene Gc element in vitro.

Transcriptional properties of proglucagon promoter sequences harboring mutations in G1. To ascertain whether the mutations that impair cdx-2/3 binding are associated with a reduction in proglucagon promoter activity in glucagon-producing cell lines, we analyzed the transcriptional properties of proglucagon-luciferase fusion genes that contained wild-type or mutant Gc sequences in their native genomic orientation. The M3 mutation (which did not significantly impair Gc2 complex formation) resulted in a small drop in proglucagon promoter activity in islet InR1-G9 cells (−74M3 [Fig. 6]). In contrast, the M6 mutation (which is clearly associated with impaired cdx-2/3 binding) resulted in a marked decrease in promoter activity in InR1-G9 islet cells. The transcriptional activity of the identical plasmids was next assessed in the two mouse intestinal cell lines. The results of the transfection experiments demonstrated that the transcriptional activity of the −74M6 plasmid was also clearly reduced in the GLUTag and STC-1 cell lines (Fig. 6), suggesting that the binding of cdx-2/3 to the Gc element is functionally important for the control of proglucagon promoter activity in cells of both islet and intestinal lineage.

DISCUSSION

The proglucagon gene proximal promoter element G1 (bp −100 to −52) is sufficient for cell-specific expression of the proglucagon gene in alpha cells (24, 27). The importance of this element for proglucagon gene transcription is emphasized by the 100% conservation of G1 sequences between the rat and human proglucagon genes (27). The data presented here represent the first identification of a specific transcription factor that mediates G1-dependent activation of proglucagon gene transcription in both islet and intestinal cell lines. Furthermore, the expression of endogenous cdx-2/3, in both small and large intestines (13, 14, 36) as well as in pancreatic islets, is

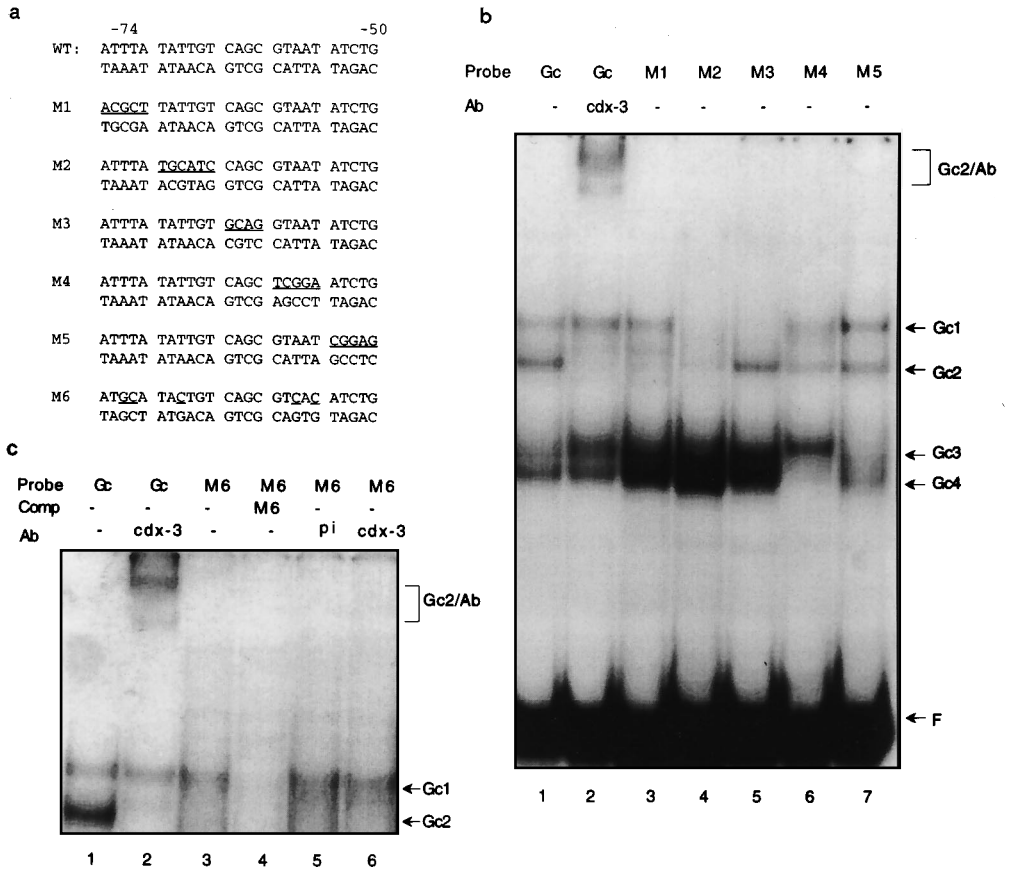


FIG. 5. Identification of Gc nucleotide subdomains important for cdx-2/3 binding. (a) The sequence of the wild-type Gc element (WT) and those of the Gc block mutations M1 through M6 are shown, with the specific nucleotide changes underlined. (b) EMSA using the M1 through M5 probes and InR1-G9 islet nuclear extracts. Nuclear extract (7.5 μg) was incubated with the wild-type or mutated M1 through M5 sequences. The migration positions of the Gc1 through Gc4 complexes are indicated. F, free probe. Ab, antisera. (c) EMSA analysis of InR1-G9 nuclear extract binding to the M6 probe. InR1-G9 nuclear extract (7.5 μg) was incubated with or without (−) cdx-3 antiserum before the addition of labeled probe GcM6 (M6). Comp, competitor sequences. Unlabeled M6 at a 200-fold molar excess was utilized (lane 5) to verify specificity of binding events. Ab, antibody; pi, preimmune. Gc1 and Gc2 denote migration positions of the respective complexes. Gc2/Ab denotes the position of the supershifted cdx-2/3 complex.

consistent with the known distribution of glucagon-producing cell types.

The mouse *cdx-2* homeobox sequence was first isolated by reverse transcription-PCR with RNA isolated from purified small intestinal crypts (14). Although the *cdx-2* homeobox sequence was 88% identical (at the amino acid level) to the previously described *cdx-1* gene (12), Southern blot analysis using the *cdx-2* sequence was consistent with the presence of a unique single copy gene, distinct from *cdx-1* (14). The subsequent isolation of the hamster *cdx-3* sequence (11), followed by the cloning of the cDNA encoding mouse cdx-2 (13, 36), demonstrated that these two sequences share 93.6% amino acid identity, consistent with cdx-3 representing the hamster homolog of mouse cdx-2. Analysis of cdx-2/3 RNA transcripts by Northern blotting and RNase protection demonstrated that cdx-2/3 is expressed in the small and large intestines in rat embryos and adult mice (13, 36). Although cdx-2/3 transcripts were not detected (by these techniques) in pancreatic RNA, this likely reflects the difficulty in detecting low-abundance RNA transcripts present in islet cells that contribute to only a small percentage of total pancreatic RNA. Furthermore, as the proglucagon-producing endocrine cell lines used in our study are likely to be heterogeneous, only a subpopulation of them may express both proglucagon and cdx-2/3, and this issue may be addressed in the future by in situ hybridization and immunocytochemistry.

Recent studies have demonstrated that mouse cdx-2 binds to and transactivates the sucrase-isomaltase (SI) gene promoter through the intestine-specific SIF1 promoter element (Fig. 7) (36). Interaction of cdx-2 with the SIF1 element may occur through either a monomer or homodimer conformation, with the extent of homodimer formation dependent on the redox state of the binding reaction (36). The redox regulation of homeodomain binding appears to be an increasingly common mode of regulating cooperative DNA binding and appears to be dependent on the presence of specific cysteine residues in homeodomain proteins (10, 34, 36). Although we detected only a single cdx-2/3 complex with the Gc probe, the failure to observe evidence for multiple cdx-2/3-containing complexes when using islet or intestinal extracts may potentially be attributable to the redox state of the binding reactions or to the unique structure of the proglucagon gene cdx-2/3 binding site.

The DNA binding specificity of caudal homeodomain proteins has been carefully examined by using cHox-cad (CdxA) and a binding site selection assay (19). This chicken homeodomain protein binds the consensus sequence A-(A/T)-T-(A/T)-A-T-(A/G) (Fig. 7). Four similar binding sites for the *Drosophila* caudal protein in the *fushi tarazu* gene promoter were also

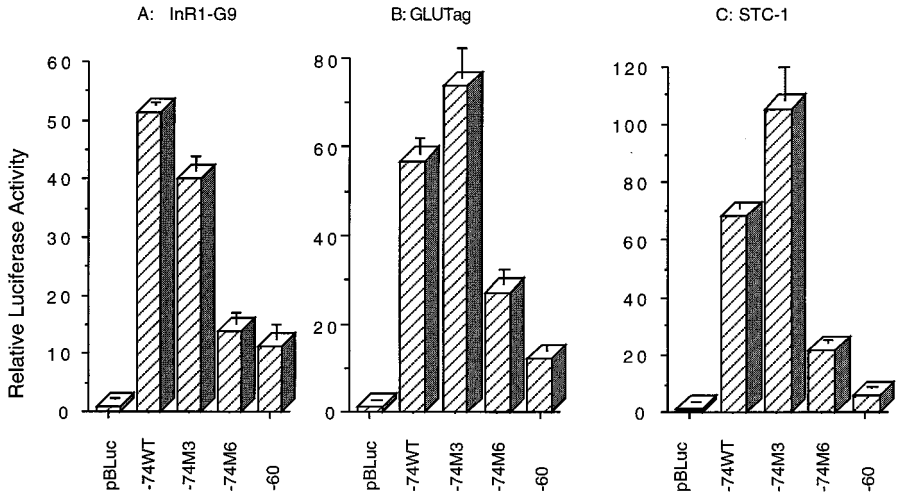


FIG. 6. Transcriptional activity of proglucagon promoter plasmids in islet and intestinal cell lines. Proglucagon-luciferase fusion genes containing wild-type proglucagon gene sequences (−74WT) or mutations in the Gc element (−74M3 and −74M6) were transfected into InR1-G9 (A), GLUTag (B), and STC-1 (C) cell lines. The data represent the mean relative luciferase activity ± standard error of the mean after normalization of the values obtained to the activity of the promoterless plasmid pBLuc in the same transfection.

previously identified (19) (Fig. 7), consistent with the observations that the caudal protein is a direct transcriptional activator of the *fushi tarazu* promoter both in *Drosophila* embryos and in transfected Schneider cells in vitro (3). Comparison of the proglucagon gene G1 (Gc) element that contains the cdx-2/3 binding sites with the cdx-2/3 binding sites identified in the SI promoter (36) reveals a number of interesting differences. The core region of the cdx-2/3 binding site in the SI promoter is an inverted repeat of ATAAA, separated by two nucleotides (Fig. 7). Each half site is capable of interacting with a single molecule of cdx-2, forming a monomer. The cdx-2 protein also binds as a dimer to the whole element, as influenced by the redox state (36). If the cdx-2/3 binding half site in SIF1 is considered to be 7 bases in length, as suggested by the cHox-cad consensus, the two SIF1 half sites correspond very well to the consensus sequence for caudal binding sites (19, 36). In contrast, the two AT-rich subdomains (which interact with cdx-2/3) in the proglucagon gene Gc element are separated by at least 7 nucleotides. Whereas the 5′ half site ATTTATA (from −73 to −67) matches the caudal consensus binding site, the 3′ half site (TAATAT) is less well conserved (Fig. 7); however, replacement of the AT-rich sequences in the 3′ half site (as seen with the M4 Gc mutant) clearly disrupts cdx-2/3 binding to the proglucagon promoter. Taken together, comparison of the data obtained from analysis of the proglucagon gene Gc element with the results of previous studies using the SI and insulin gene promoters (11, 36) suggests that cdx-2/3 is capable of recognizing significantly diverse DNA target sequences. The functional relevance of the nucleotide differences in cdx-2/3 target sequences for cdx-2/3 binding, as well as the importance of the differences in spacing between the two half sites, remains to be determined.

```
CDXA binding site:          A   A/T   T   A/T   A   T   A/G

Ftz promoter          a)    T     T   T     T   A   G     G
                      b)    G     T   T     T   A   T     G
                      c)    T     T   T     T   A   T     G
                      d)    C     T   T     T   A   T     G
```

```
FLAT element in the rat insulin I gene minienhancer

T G T T A A T A A T C T A A T T A C C C
A C A A T T A T T A G A T T A A T G G G
  FLAT F                FLAT E
```

```
SIF1 element in the sucrase isomaltase promoter

G G G T G C A A T A A A A C T T T A T G A G T A
C C C A C G T T A T T T T G A A A T A C T C A T
```

```
Gc element in the rat and human proglucagon promoter

A T T T A T A T T G T C A G C G T A A T A T C T G
T A A A T A T A A C A G T C G C A T T A T A G A C
```

FIG. 7. Comparison of the DNA consensus sequences for binding of caudal homeodomain proteins. FLAT, far-linked AT rich.

The identification of cdx-2/3 as a major component of the proglucagon gene Gc2 complex provides an important first step towards the identification of the specific binding proteins that regulate proglucagon promoter specificity in islet and intestinal cells through the G1 element. Nevertheless, we consistently observed a number of additional distinct complexes, for example, Gc1, Gc3, and Gc4, with the proglucagon gene Ga, Gb, and Gc probes and both islet and intestinal extracts. Furthermore, although cdx-2/3 constitutes a major component of the Gc2 complex detected with the Gc probe, the Gc2-like complex formed with the Ga and Gb probes, although of similar mobility, did not appear to contain cdx-2/3, as observed in EMSA experiments using specific cdx-2/3 antisera. These observations are consistent with the existence of additional distinct G1-binding proteins that may also contribute to the regulation of cell-specific promoter activity through the G1 element. The identification of these proteins, and their role in controlling proglucagon gene transcription in islet and intestinal cells, as well as analysis of the functional role of cdx-2/3 in vivo, should further our understanding of the molecular control of cell-specific gene expression in the endocrine cells of the islets and intestine.

ACKNOWLEDGMENTS

T.J. is supported by a Fellowship Award from the Ontario Ministry of Health. D.J.D. is a Scientist of the Medical Research Council of Canada (MRC). This work was supported by an operating grant from the MRC.

We thank Feng Wang for expert technical assistance.

REFERENCES

1. **Bell, G. I., R. F. Santerre, and G. T. Mullenbach.** 1983. Hamster preproglucagon contains the sequence of glucagon and two related peptides. Nature (London) **302:**716–718.
2. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Chem. **162:**156–159.
3. **Dearolf, C. R., J. Topol, and C. S. Parker.** 1989. The caudal gene product is a direct activator of *fushi tarazu* transcription during *Drosophila* embryogenesis. Nature (London) **341:**340–343.
4. **DeWet, J. R., K. V. Wood, M. Deluca, D. R. Helinski, and S. Subramani.** 1987. Firefly luciferase gene: structure and expression in mammalian cells. Genes Dev. **8:**1–13.
5. **Drucker, D. J., and S. Asa.** 1988. Glucagon gene expression in vertebrate brain. J. Biol. Chem. **263:**13475–13478.
6. **Drucker, D. J., and P. L. Brubaker.** 1989. Proglucagon gene expression is regulated by a cyclic AMP-dependent pathway in rat intestine. Proc. Natl. Acad. Sci. USA **86:**3953–3957.
7. **Drucker, D. J., T. Jin, S. L. Asa, T. A. Young, and P. L. Brubaker.** 1994. Activation of proglucagon gene transcription by protein kinase A in a novel mouse enteroendocrine cell line. Mol. Endocrinol. **8:**1646–1655.
8. **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.
9. **Drucker, D. J., J. Philippe, and S. Mojsov.** 1988. Proglucagon gene expression and posttranslational processing in a hamster islet cell line. Endocrinology **123:**1861–1867.
10. **Galang, C. K., and C. A. Hauser.** 1993. Cooperative DNA binding of the human HoxB5 (Hox2.1) protein is under redox regulation in vitro. Mol. Cell. Biol. **13:**4609–4617.
11. **German, M. S., J. Wang, R. B. Chadwick, and W. J. Rutter.** 1992. Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. Genes Dev. **6:**2165–2176.
12. **Hu, Y., J. Kazenwadel, and R. James.** 1993. Isolation and characterization of the murine homeobox gene *CDX-1*. J. Biol. Chem. **268:**27214–27225.
13. **James, R., T. Erler, and J. Kazenwadel.** 1994. Structure of the murine homeobox gene cdx-2. J. Biol. Chem. **269:**15229–15237.
14. **James, R., and J. Kazenwadel.** 1991. Homeobox gene expression in the intestinal epithelium of adult mice. J. Biol. Chem. **266:**3246–3251.
15. **Knepel, W., L. Jepeal, and J. F. Habener.** 1990. A pancreatic islet cell-specific enhancer-like element in the glucagon gene contains two domains binding distinct cellular proteins. J. Biol. Chem. **265:**8725–8735.
16. **Knepel, W., M. Vallejo, J. A. Chafitz, and J. F. Habener.** 1991. The pancreatic islet-specific glucagon G3 transcription factors recognize control elements in the rat somatostatin and insulin-I genes. Mol. Endocrinol. **5:**1457–1466.
17. **Lee, Y. C., S. L. Asa, and D. J. Drucker.** 1992. Glucagon gene 5′-flanking sequences direct expression of SV40 large T antigen to the intestine producing carcinoma of the large bowel in transgenic mice. J. Biol. Chem. **267:**10705–10708.
18. **Leonard, J., B. Peers, T. Johnson, K. Ferrei, S. Lee, and M. R. Montminy.** 1993. Characterization of somatostatin transactivating factor-1, a novel homeobox factor that stimulates somatostatin gene expression in pancreatic islet cells. Mol. Endocrinol. **7:**1275–1283.
19. **Margalit, Y., S. Yarus, E. Shapira, Y. Gruenbaum, and A. Fainsod.** 1993. Isolation and characterization of target sequences of the chicken CdxA homeobox gene. Nucleic Acids Res. **21:**4915–4922.
20. **Miller, C. P., Jr., R. E. McGehee, and J. F. Habener.** 1994. IDX-1: a new homeodomain transcription factor expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene. EMBO J. **13:**1145–1156.
21. **Mlodzik, M., A. Fjose, and W. A. Gehring.** 1985. Isolation of *caudal*, a *Drosophila* homeo box-containing protein with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage. EMBO J. **4:**2961–2969.
22. **Mlodzik, M., and W. J. Gehring.** 1987. Expression of the *caudal* gene in the germ line of *Drosophila*: formation of an RNA and protein gradient during early embryogenesis. Cell **48:**465–478.
23. **Mojsov, S., G. Heinrich, I. B. Wilson, M. Ravazzola, L. Orci, and J. F. Habener.** 1986. Preproglucagon gene expression in pancreas and intestine diversifies at the level of post-translational processing. J. Biol. Chem. **261:**11880–11889.
24. **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.
25. **Orskov, C., J. J. Holst, S. S. Poulsen, and P. Kirkegaard.** 1987. Pancreatic and intestinal processing of proglucagon in man. Diabetologia **30:**874–881.
26. **Peers, B., J. Leonard, S. Sharma, G. Teitelman, and M. R. Montminy.** 1994. Insulin expression in pancreatic islet cells relies on cooperative interactions between the helix loop helix factor E47 and the homeobox factor STF-1. Mol. Endocrinol. **8:**1798–1806.
27. **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 DNA-binding proteins. Mol. Cell. Biol. **8:**4877–4888.
28. **Philippe, J., S. Mojsov, D. J. Drucker, and J. F. Habener.** 1986. Proglucagon processing in rat islet cell line resembles phenotype of intestine rather than pancreas. Endocrinology **119:**2833–2839.
29. **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.
30. **Philippe, J., C. Morel, and V. R. Prezioso.** 1994. Glucagon gene expression is negatively regulated by hepatocyte nuclear factor 3β. Mol. Cell. Biol. **14:**3514–3523.
31. **Powers, A. C., F. Tedeschi, K. E. Wright, J. S. Chan, and J. F. Habener.** 1989. Somatostatin gene expression in pancreatic islet cells is directed by cell-specific DNA control elements and DNA-binding proteins. J. Biol. Chem. **264:**10048–10056.
32. **Rindi, G., S. G. N. Grant, Y. Yiangou, M. A. Ghatei, S. R. Bloom, V. L. Bautch, E. Solcia, and J. M. Polak.** 1990. Development of neuroendocrine tumors in the gastrointestinal tract of transgenic mice. Am. J. Pathol. **136:**1349–1363.
33. **Rudnick, A., T. Y. Ling, H. Odagiri, W. J. Rutter, and M. S. German.** 1994. Pancreatic beta cells express a diverse set of homeobox genes. Proc. Natl. Acad. Sci. USA **91:**12203–12207.
34. **Sanchez-Garcia, I., and T. H. Rabbitts.** 1993. Redox regulation of in vitro DNA-binding activity by the homeodomain of the isl-1 protein. J. Mol. Biol. **231:**945–949.
35. **Schreiber, E., P. Matthias, M. M. Muller, and W. Schaffner.** 1989. Rapid detection of octamer binding proteins with 'mini-extracts' prepared from a small number of cells. Nucleic Acids Res. **17:**6419.
36. **Suh, E., L. Chen, J. Taylor, and P. G. Traber.** 1994. A homeodomain protein related to caudal regulates intestine-specific gene transcription. Mol. Cell. Biol. **14:**7340–7351.