The EMBO Journal vol.13 no.5 pp.1145-1156, 1994

IDX-1: a new homeodomain transcription factor expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene

Christopher P.Miller, Robert E.McGehee,Jr[1] and Joel F.Habener

Laboratory of Molecular Endocrinology, Massachusetts General Hospital and Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02114, USA
[1]Present address: Department of Pediatric Research, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

Communicated by R.A.Flavell

We describe the cloning from a rat islet somatostatin-producing cell line of a 1.4 kb cDNA encoding a new homeoprotein, IDX-1 (islet/duodenum homeobox-1), with close sequence similarity to the *Drosophila melanogaster* homeobox protein Antennapedia (Antp) and the *Xenopus laevis* endoderm-specific homeoprotein XlHbox8. Analyses of IDX-1 mRNA and protein in rat tissues show that IDX-1 is expressed in pancreatic islets and ducts and in the duodenum. In electrophoretic mobility shift assays IDX-1 binds to three sites in the 5′ flanking region of the rat somatostatin gene. In co-transfection experiments IDX-1 transactivates reporter constructs containing somatostatin promoter sequences, and mutation of the IDX-1 binding sites attenuates transactivation. Reverse transcription-polymerase chain reaction of islet RNA using degenerate amplimers for mRNAs encoding homeoproteins indicates that IDX-1 is the most abundant of 12 different Antp-like homeodomain mRNAs expressed in adult rat islets. The pattern of expression, relative abundance and transcriptional regulatory activity suggests that IDX-1 may be involved in the regulation of islet hormone genes and in cellular differentiation in the endocrine pancreas and the duodenum.

Key words: DNA binding/homeodomain/pancreatic islet/somatostatin gene regulation//transcriptional activation

Introduction

During embryonic development in rodents, the primordial pancreatic anlaga are formed from outpocketings of the primitive gut on or around gestation day 10 (Wessells and Evans, 1968; Pictet and Rutter, 1972). Cells within the pancreatic anlaga subsequently differentiate and give rise to exocrine and endocrine tissues. The initial formation of the pancreatic anlaga coincides with the first appearance of digestive enzymes of the exocrine pancreas and the hormones produced by the endocrine pancreas (Pictet and Rutter, 1972; Han *et al.*, 1986). In the early stages of pancreatic development in rodents the endocrine cells are found in close proximity to the pancreatic ducts, and are believed to be derived from stem cell populations in or around the ducts (Yoshinari and Diakoku, 1982; Teitelman and Lee, 1987; Alpert *et al.*, 1988; Dudek *et al.*, 1991). During pancreatic islet development there is an ordered progression in the appearance of the major islet hormones. Glucagon-producing α-cells are first detected on gestation day 10, whereas insulin-producing β-cells and somatostatin-producing δ-cells first appear on gestation days 12 and 17, respectively (Pictet and Rutter, 1972; Teitelman and Lee, 1987). Cellular phenotypes are established by selective transcriptional activation or repression of cell-specific sets of genes, which are accomplished by the interaction of DNA binding proteins with their corresponding DNA regulatory elements and with the proteins of the general transcription machinery. Homeodomain transcription factors are sequence-specific DNA binding proteins known to be important in directing embryonic development and determining differentiated cell identity in a wide variety of organisms (Gehring, 1987; Holland and Hogan, 1988; Blumberg *et al.*, 1991; Bürglin *et al.*, 1989; Nohno *et al.*, 1991; Bellmann and Werr, 1992; Renucci *et al.*, 1992; Schummer *et al.*, 1992; Levine and Schechter, 1993). The homeodomain is a highly conserved, 61 amino acid, DNA binding structure first identified in homeotic selector genes of *Drosophila melanogaster* (Gehring, 1987; Scott *et al.*, 1989). As part of an ongoing effort to identify and characterize DNA binding proteins involved in pancreatic islet development and differentiation, we sought to clone homeodomain-encoding cDNAs from a λgt11 cDNA library prepared from the somatostatin-producing islet cell line RIN1027-B2 (Philippe *et al.*, 1987).

The homeodomain consists of three α-helical regions separated by a loop and a turn. The high degree of conservation of amino acid sequences within portions of the first and third helices (Scott *et al.*, 1989) has facilitated the cloning of homeodomain cDNAs (Bürglin *et al.*, 1989; James and Kazenwadel, 1991; Murtha *et al.*, 1991; Singh *et al.*, 1991; Nazarali *et al.*, 1992; Patel *et al.*, 1992; Levine and Schechter, 1993). The most highly conserved region of the homeodomain is the amino acid sequence KIWFQN within the DNA recognition helix (helix 3). We therefore used a degenerate oligonucleotide corresponding to this sequence (Bürglin *et al.*, 1989) to screen the RIN1027-B2 cDNA library for homeodomain-containing cDNAs. A second highly conserved amino acid sequence, ELEKEF, within the first helix of the homeodomain was utilized to design a degenerate polymerase chain reaction (PCR) amplimer (James and Kazenwadel, 1991). This was used, in conjunction with a degenerate PCR amplimer corresponding to the KIWFQN motif in the third helix, to amplify homeodomain-containing cDNAs from rat islet cDNA. By using this combined approach, we isolated a complete cDNA for IDX-1 (islet/duodenum homeobox-1), a new homeodomain protein from the RIN1027-B2 cDNA library, and surveyed homeodomain proteins expressed in rat islets. Experiments presented here describe the cloning of IDX-1, its cell- and tissue-specific expression in pancreatic islets and ducts, and in duodenum, and its DNA binding and transcriptional regulatory activities. Furthermore, we show that IDX-1 appears to be the most abundant homeodomain

mRNA in rat islets, and we identify 11 additional homeodomain mRNAs expressed in rat islets.

Results

Isolation and sequence analysis of the IDX-1 cDNA

A RIN1027-B2 cDNA library was screened with a degenerate oligonucleotide corresponding to the highly conserved KIWFQN amino acid sequence in the third helix of the Antennapedia (Antp) homeodomain (Bürglin *et al.*, 1989), resulting in the isolation of the 1.4 kb IDX-1 cDNA (Figure 1A). The IDX-1 cDNA contains an open reading frame encoding the 283 amino acid protein, 150 nucleotides of 5′ untranslated sequence and 500 nucleotides of 3′ untranslated sequence. The IDX-1 protein has an estimated molecular weight of 31 kDa, contains a homeodomain and has proline-rich regions in its amino- and carboxy-terminal regions (Figure 1A and B). As expected, the IDX-1 homeodomain is similar to Antp, particularly in the first and third helices (Figure 1C). These regions are the most highly conserved among the vast majority of known homeoproteins (Scott *et al.*, 1989). Immediately below IDX-1 in Figure 1C is the amino acid sequence for a portion of the homeodomain of XlHbox8, a homeobox protein expressed in the developing and mature duodenum and pancreas of *Xenopus laevis*. The only published sequence for XlHbox8 is from a partial clone beginning at amino acid residue 19 in the homeodomain (Wright *et al.*, 1988). There is no available XlHbox8 sequence 5′ to this site. The amino acid sequence of IDX-1 is identical to that of XlHbox8 from residue 19 to residue 60 of the homeodomain, while the sequences of the two proteins diverge completely just carboxy-proximal to the homeodomain. There is 80% nucleic acid similarity between IDX-1 and XlHbox8 within the homeodomain. The high degree of sequence similarity in the IDX-1 and XlHbox8 homeodomains and similar patterns of expression suggests that the two proteins may be closely related, or that IDX-1 is the rat homologue of XlHbox8. The amino acid sequences of IDX-1 and XlHbox8 are also highly similar to Htr-A2, a homeodomain-containing gene isolated from the leech, *Helobdella triseralis*.

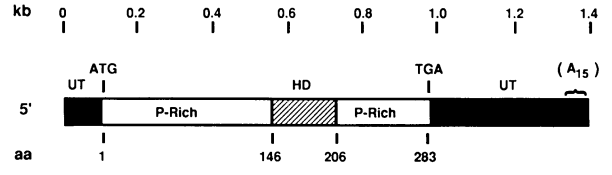
Tissue distribution of IDX-1 expression

Northern analysis of total RNA from adult rat tissues indicates that the 1.5 kb IDX-1 mRNA is expressed in pancreatic islets and duodenum (Figure 2A–C). Initially, a multiple tissue Northern blot showed that IDX-1 mRNA was detected only in intestine (Figure 2A). This observation prompted us to examine other segments of the rat gastrointestinal tract for IDX-1 expression. As shown in Figure 2B, IDX-1 mRNA can be detected in the duodenum, but not in any other section of the gastrointestinal tract. IDX-1 mRNA is detected at uniform levels along the first 8 cm of the rat duodenum (data not shown). Since IDX-1 was cloned from a rat pancreatic islet-derived cell line, it was of interest to determine whether IDX-1 mRNA is

A

```
GAATTCCGGGGCGCTGAGAGTCCGTGAGCTGCCCAGCGCCTAAGGCCTGGCTTGTAGCTCC     61

                                  M  N  S  E  E  Q  Y  Y      8
CTACCCCGGGCTGCCGGCCCCGAAGTGCCGGCTGCCACCATGAATAGTGAGGAGCAGTACT    122

  A  A  T  Q  L  Y  K  D  P  C  A  F  Q  R  G  P  V  P  E  F     28
ACGCGGCCACACAGCTCTACAAGGACCCGTGCGCATTCCAGAGGGGTCCGGTGCCAGAGTT    183

  S  A  N  P  P  A  C  L  Y  M  G  R  Q  P  P  P  P  P  T  P     49
CAGTGCTAATCCCCCTGCGTGCCTGTACATGGGCCGCCAGCCCCCACCTCCGCCGACACCC    244

  Q  F  A  G  S  L  G  T  L  E  Q  G  S  P  P  D  I  S  P  Y  E     69
CAGTTTGCAGGCTCGCTGGGAACGCTGGAACAGGGAAGTCCCCCGGACATCTCCCCATACG    305

  V  P  P  L  A  D  D  P  A  G  A  H  L  H  H  H  L  P  A  Q     89
AAGTGCCCCCGCTCGCCGATGACCCGGCTGGCGCGCACCTCCACCACCACCTCCCAGCTCA    366

  L  G  L  A  H  P  P  P  G  P  F  P  N  G  T  E  T  G  G  L     110
GCTCGGGCTCGCCCATCCACCTCCCGGACCTTTCCCGAATGGAACCGAGACTGGGGGCCTG    427

  E  E  P  S  R  V  H  L  P  F  P  W  M  K  S  T  K  A  H  A  W     130
GAAGAGCCCAGCCGCGTTCATCTCCCTTTCCCGTGGATGAAATCCACCAAAGCTCACGCGT    488

  K  S  Q  W  A  G  G  A  Y  A  A  E  P  E  E  N  K  R  T  R     150
GGAAAAGCCAGTGGGCAGGAGGTGCATACGCAGCAGAACCGGAGGAGAATAAGAGGACCCG    549

  T  A  Y  T  R  A  Q  L  L  E  L  E  K  E  F  L  F  N  K  Y     171
TACAGCCTACACTCGGGCCCAGCTGCTGGAGCTGGAGAAGGAATTCTTATTTAACAAATAC    610

  I  S  R  P  R  R  V  E  L  A  V  M  L  N  L  T  E  R  H  I  K     191
ATCTCCCGGCCTCGCCGGGTGGAGCTGGCAGTGATGCTCAACTTGACTGAGAGACACATCA    671

  I  W  F  Q  N  R  R  M  K  W  K  K  E  E  D  K  K  R  S  S     211
AAATCTGGTTCCAAAACCGTCGCATGAAGTGGAAGAAAGAGGAAGATAAGAAACGTAGTAG    732

  G  T  T  S  G  G  G  G  G  E  E  P  E  Q  D  C  A  V  T  S     232
CGGGACAACGAGCGGGGGCGGTGGGGGCGAAGAGCCGGAGCAGGATTGTGCCGTAACCTCG    793

  G  E  E  L  L  A  L  P  K  P  P  P  P  G  G  V  V  P  S  G  V     252
GGCGAGGAGCTGCTGGCATTGCCAAAGCCACCACCTCCCGGAGGTGTTGTGCCCTCAGGCG    854

  P  A  A  A  R  E  G  R  L  P  S  G  L  S  A  S  P  Q  P  S     272
TCCCTGCTGCTGCCCGGGAGGGCCGACTGCCTTCCGGCCTTAGTGCGTCCCCACAGCCCTC    915

  S  I  A  P  L  R  P  Q  E  P  R  *                          283
CAGCATCGCGCCACTGCGACCGCAGGAACCCCGGTGAGGACCGCAGGCTGAGGGTGAGCGG    976

GTCTGGGACCCAGAGTGCGGACATGGGCATGGGCCCGGGCAGCTGGATAAGGGAGGGGATC   1037
ATGAGGCTTAACCTAAACGCCACACACAAGGAGAACATTCTTCTTGGGGGCACAAGAGCCA   1098
GTTGGGTATACCAGCGAGATGCTGGCAGACCTCTGGGAAAAAAAAAGACCCGAGCTTCTGA   1159
AAACTTTGAGGCTGCCTCTCGTGCCATGTGAACCGCCAGGTCTGCCTCTGGGACTCTTTCC   1220
TGGGACCAATTTAGAGAATCAGGCTCCCAACTGAGGACAATGAAAAGGTTACAAACTTGAG   1281
CGGTCCCATAACAGCCACCAGGCGAGCTGGACCGGGTGCCTTTGACTGGTCGCCGAGCAA   1342
TCTAAGGTTGAGAATAAAGGGAGCTGTTTGAGGTTTCAAAAAAAAAAAAAACCGGAATTC   1404
```

B



C

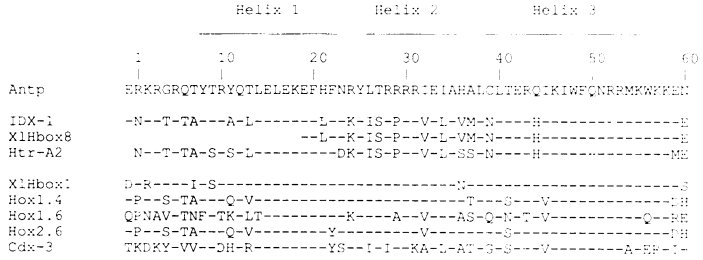


Fig. 1. (**A**) Sequence of the IDX-1 cDNA and encoded protein. The homeodomain is underlined. (**B**) Schematic representation of the IDX-1 cDNA and encoded protein. Translation start (ATG) and stop (TGA) codons are indicated. HD, homeodomain; P-Rich, proline-rich region; UT, untranslated region. (**C**) Comparison of the amino acid sequences in the homeodomains of Antp, IDX-1 and selected other homeodomain proteins. Numbers above the Antp sequence indicate amino acid position within the homeodomain. Dashes denote sequence identity with Antp. Note that the only available sequence for XlHbox8 is from a partial clone beginning at amino acid residue 20 in the homeodomain (Wright *et al.*, 1988).

expressed in rat islets. Northern analysis indicates that the IDX-1 mRNA is detected in rat islet RNA (Figure 2C). Also shown in Figure 2C is a weakly hybridizing band from mouse duodenum RNA that is ~2.0 kb long. Lack of detectable IDX-1 mRNA in RNA preparations from whole pancreas is attributed to the fact that islets represent only 1–2% of the whole pancreas and that some islet-specific mRNAs are diluted out to undetectable levels by the non-islet contribution to total pancreatic RNA. IDX-1 mRNA is also detected in RNA prepared from fetal rat whole intestine at gestational day 20 (Figure 2D). During embryonic development, the pancreas is derived from the duodenum. The observation that IDX-1 mRNA is detected in adult rat islets and duodenum and in fetal intestine, suggests a role for IDX-1 in pancreatic development.

Western immunoblot analysis of rat tissue extracts indicates that, as for IDX-1 mRNA, IDX-1 protein is detected in rat islets and duodenum. The major protein displaying IDX-1 immunoreactivity in Western immunoblot analysis of rat islet whole cell extracts, RIN1027-B2 nuclear and whole cell extracts, and crude nuclear lysates from duodenum migrates with an apparent molecular weight of 42 kDa (Figure 2E). COS-1 cells transfected with an IDX-1 eukaryotic expression vector also produce IDX-1 protein with an apparent molecular weight of 42 kDa. In addition, the major bands of IDX-1 immunoreactivity and radioactivity from IDX-1 transcribed and translated *in vitro*, or immunoprecipitated from ^{35}S-labelled RIN1027-B2 cells are also 42 kDa (data not shown). No IDX-1 protein is detected in jejunum crude nuclear lysates. As a control, CREB protein is easily detected in all of these extracts and lysates, indicating that lack of IDX-1 protein in jejunum was not due to sample degradation or variable loading.

Expression of IDX-1 in cultured cell lines

Several islet and non-islet cell lines were assayed for IDX-1 expression by Northern RNA blot and Western immunoblot (Figure 3A and B). RIN1046-38 and RIN1027-B2 cells are

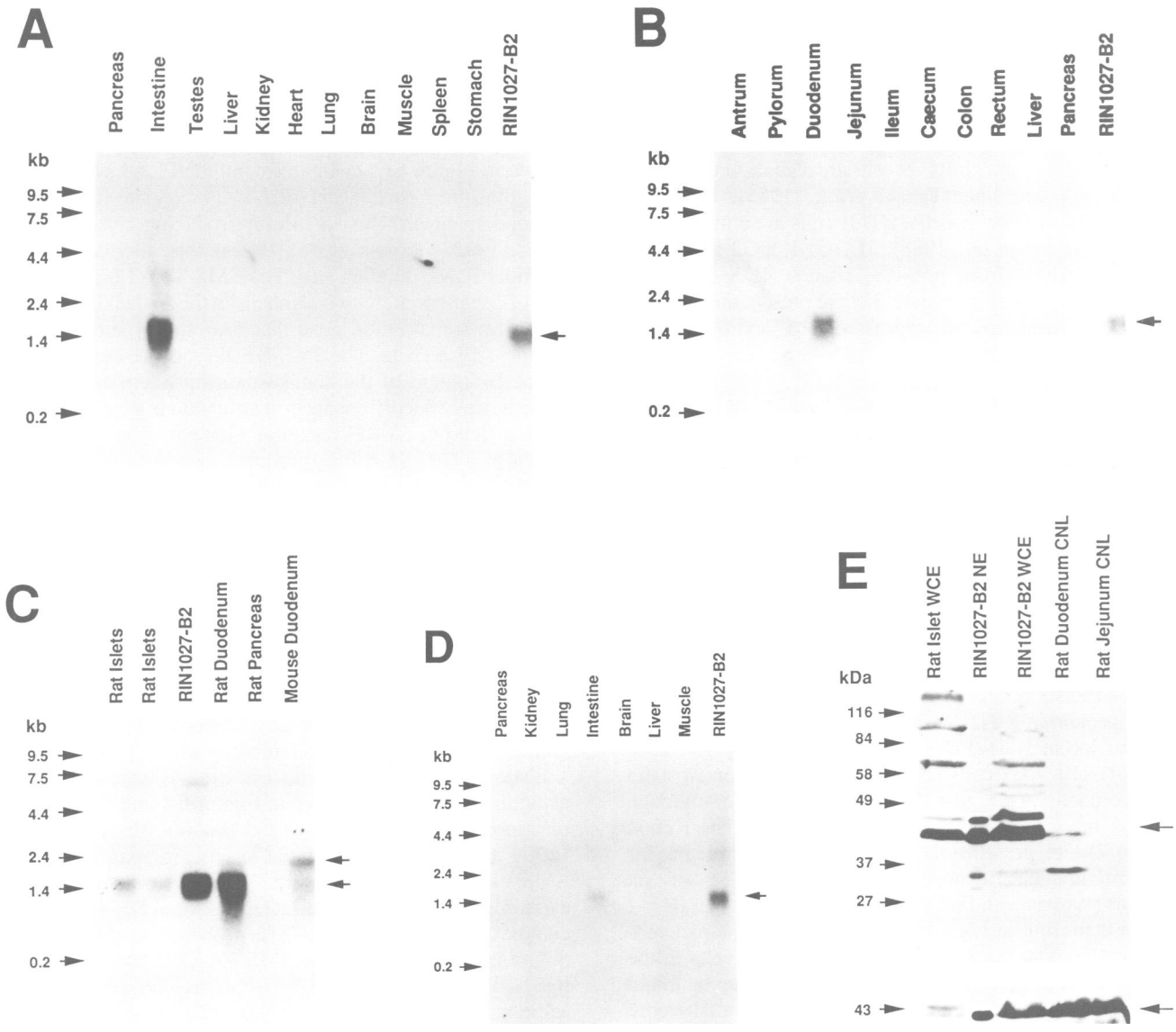


Fig. 2. IDX-1 is expressed in pancreatic islets and duodenum. (**A**–**C**) Northern analyses of IDX-1 mRNA in adult rat tissues. Twenty micrograms of total RNA from each tissue or 10 μg of total cellular RNA from RIN1027-B2 cells were loaded per lane. (**D**) Northern analysis of IDX-1 mRNA in fetal rat tissues (gestational day 20). (**E**) Western immunoblot analysis of IDX-1 (top) and CREB (bottom) proteins in adult rat tissues. (NE, nuclear extracts; WCE, whole cell extracts; CNL, crude nuclear lysates.)

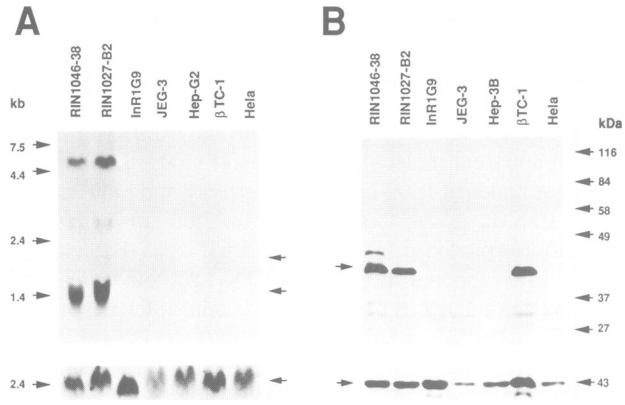


Fig. 3. IDX-1 is expressed in several islet-derived cell lines but is not detected in non-islet cell lines. (**A**) Northern analysis of IDX-1 (top) and actin (bottom) mRNAs in assorted cultured cell lines. Fifteen micrograms of total cellular RNA were loaded per lane. (**B**) Western immunoblot analysis of IDX-1 (top) and CREB (bottom) proteins in nuclear extracts prepared from assorted cultured cell lines.

derived from a radiation-induced rat insulinoma (Chick *et al.* 1977; Philippe *et al.*, 1987). RIN1046-38 cells produce insulin and most closely resemble islet β-cells, whereas RIN1027-B2 cells produce somatostatin and have more features of islet δ-cells. InR1G9 cells are a glucagon-secreting hamster islet cell line (Takaki *et al.*, 1986), and have characteristics of islet α-cells. βTC1 cells are a mouse islet β-cell line (Efrat *et al.*, 1988). JEG-3, HepG2 and HeLa cells are derived from non-islet tissues. IDX-1 mRNA (Figure 3A) and protein (Figure 3B) are present only in islet-derived cell lines, and are expressed in β- and δ-cell lines but not in α-cell lines.

Immunohistochemistry and in situ hybridization

Western immunoblot analysis indicated that IDX-1 protein was present in rat islets and duodenum (Figure 2E). Immunohistochemistry was performed with RIN1027-B2 cells and rat islets and duodenum to gain further information regarding the tissue and cellular localization of IDX-1 protein. Immunohistochemistry with antiserum to IDX-1 clearly shows the presence of IDX-1 in the nuclei of RIN1027-B2 cells (Figure 4B). This immunostaining is absent in cells incubated with preimmune serum (Figure 4A). Nuclear localization for IDX-1 is consistent with a proposed role as a transcriptional regulator, and with the detection of IDX-1 protein in RIN1027-B2 nuclear extracts by Western immunoblot analysis (Figures 2E and 3B). Only 10–20% of the RIN1027-B2 cells show high levels of somatostatin immunoreactivity (Figure 4C), suggesting that although IDX-1 may regulate somatostatin gene transcription (Figure 7), its presence is not the sole determinant for somatostatin gene expression. Immunostaining of rat pancreas sections with IDX-1 antiserum indicates that IDX-1 protein is present in islets but not in the exocrine pancreas (Figure 4E). The presence of immunopositive cells in the islet core and periphery suggests that IDX-1 may be found in multiple islet cell phenotypes. In rodents, the islet core contains mainly insulin-producing β-cells, whereas the islet periphery is predominantly composed of somatostatin-producing δ-cells and glucagon-producing α-cells (Orci, 1982). Within the duodenum, IDX-1 immunopositive nuclei are visible in cells of the intestinal crypts (Figure 4H and I).

In situ hybridization was performed to examine tissue and cellular localization of IDX-1 mRNA in rat pancreas and duodenum sections (Figure 5). In the pancreas, IDX-1 mRNA is most easily detected in pancreatic ducts and in mesenchymal cells surrounding the ducts (Figure 5C, and 'D' in Figure 5B). Pancreatic islets are only weakly positive for IDX-1 mRNA (Figure 5B). Within the duodenum, IDX-1 mRNA is localized to a broad band within the submucosal region (Figure 5E), and to a lesser extent, in cells at the base of the crypts (Figure 5F).

Sequence-specific DNA binding by IDX-1

Electrophoretic mobility shift assays (EMSA) were used to characterize the DNA binding properties of IDX-1. EMSA/PCR-based selection and amplification of DNA binding sites (Sun and Baltimore, 1991) indicated that bacterially expressed IDX-1 preferentially binds to the sequence 5′-TAAT(T/G)-3′ (not shown). The preference for this motif is shared by other homeoproteins (Desplan *et al.*, 1988; Scott *et al.*, 1989; Ekker *et al.*, 1991). Since the IDX-1 cDNA was isolated from a rat somatostatin-producing, islet-derived cell line (RIN1027-B2), we examined the 5′ flanking region of the rat somatostatin gene for potential binding sites. The 5′-TAAT(T/G)-3′ motif occurs three times within 500 nucleotides upstream of the transcription start site in this gene. These sequences are shown in Figure 6A, compared with the FLAT element of the rat insulin 1 gene (INS1-FLAT) a target site for several homeodomain proteins in insulin-producing cells (Emens *et al.*, 1992; German *et al.*, 1992a) that also contains a potential IDX-1 binding site. The SMS-TAAT1 and SMS-TAAT2 sequences are two elements in the 5′ flanking region of the rat somatostatin gene that bear strong similarity to the INS1-FLAT site. The SMS-UE-B sequence corresponds to the B-domain of the somatostatin upstream element, a bipartite cell-specific enhancer that functions synergistically with a nearby cAMP response element (Vallejo *et al.*, 1992a,b). The SMS-PS (proximal silencer; M.Vallejo, C.P.Miller, W.Beckman and J.K.Habener, in preparation) probe contains a 5′-TAATC-3′ sequence (non-coding strand) and was included for comparison as an element that did not contain the preferred IDX-1 target sequence. In EMSA with ^{32}P-labelled probes corresponding to these sequences, bacterially expressed IDX-1 binds to the probes containing the 5′-TAAT(T/G)-3′ motif, but does not bind to the 5′-TAATC-3′ SMS-PS probe (Figure 6B). The SMS-TAAT1 and INS1-FLAT probes form several complexes with IDX-1, which may correspond to single or multiple molecules of IDX-1 binding to the probes.

EMSA with RIN1027-B2 nuclear extracts were conducted using the same probes shown in Figure 6A and B. Binding reactions were performed with or without the addition of IDX-1 antiserum or preimmune serum. Comparison of the EMSA patterns and results from cross-competition experiments (not shown) indicate that several DNA–protein complexes are shared among these probes. One of the fastest migrating complexes for the SMS-TAAT1, SMS-TAAT2, INS1-FLAT and SMS-UE-B probes is disrupted by the addition of IDX-1 antiserum (Figure 6C), indicating IDX-1 binding to these sites. Addition of preimmune serum does not alter the EMSA pattern. A second band of lower mobility with the SMS-TAAT2 and INS1-FLAT probes is also disrupted by IDX-1 antiserum. The faster migrating band is not seen for the SMS-PS probe, and no bands are perturbed

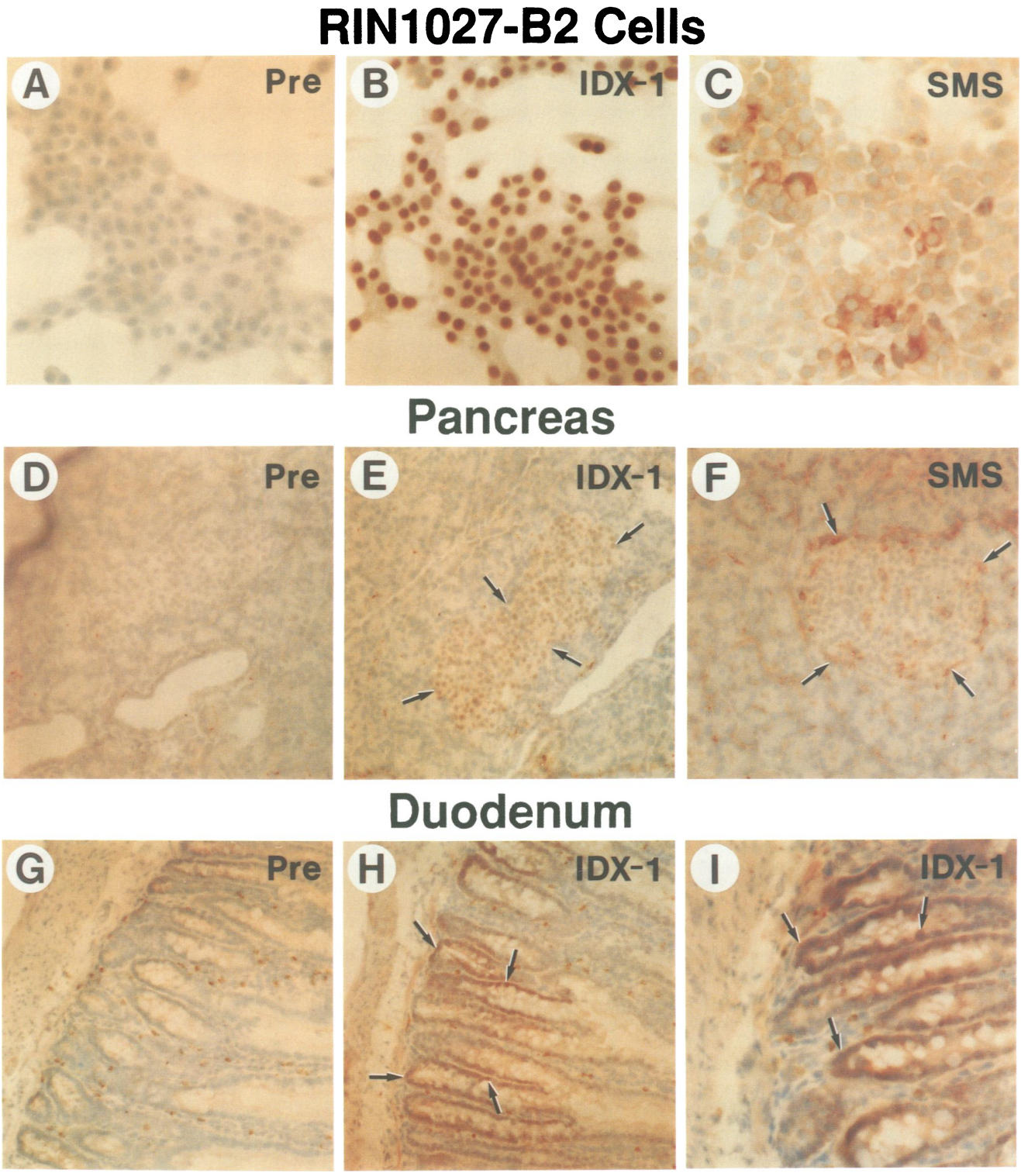


Fig. 4. Immunohistochemistry of IDX-1 or somatostatin (SMS) proteins in RIN1027-B2 cells (panels A–C), rat pancreas (D–F) and rat duodenum (G–I). Panels A, D and G: preimmune serum; panels B, E, H and I: IDX-1 antiserum; panels C and F: somatostatin antiserum. (A–C and I, ×1480; D–H, ×740.)

by the IDX-1 antiserum, consistent with the lack of SMS-PS binding by bacterially expressed IDX-1 (Figure 6B).

Transcriptional activation by IDX-1

Transcriptional activation by IDX-1 was assayed by cotransfection experiments in NIH-3T3 cells with an IDX-1 expression vector (IDX-1/pBJ5) and reporter constructs containing portions of the rat somatostatin 5′ flanking region in the plasmid pOCAT (Figure 7B and C), or multimerized IDX-1 binding sites inserted into the plasmid SMS65CAT (Figure 7D) (Powers *et al.*, 1989). The SMS900CAT reporter (Figure 7A) consists of nucleotides −900 to +54 of the rat somatostatin gene linked 5′ to the gene encoding the bacterial enzyme chloramphenicol acetyl transferase

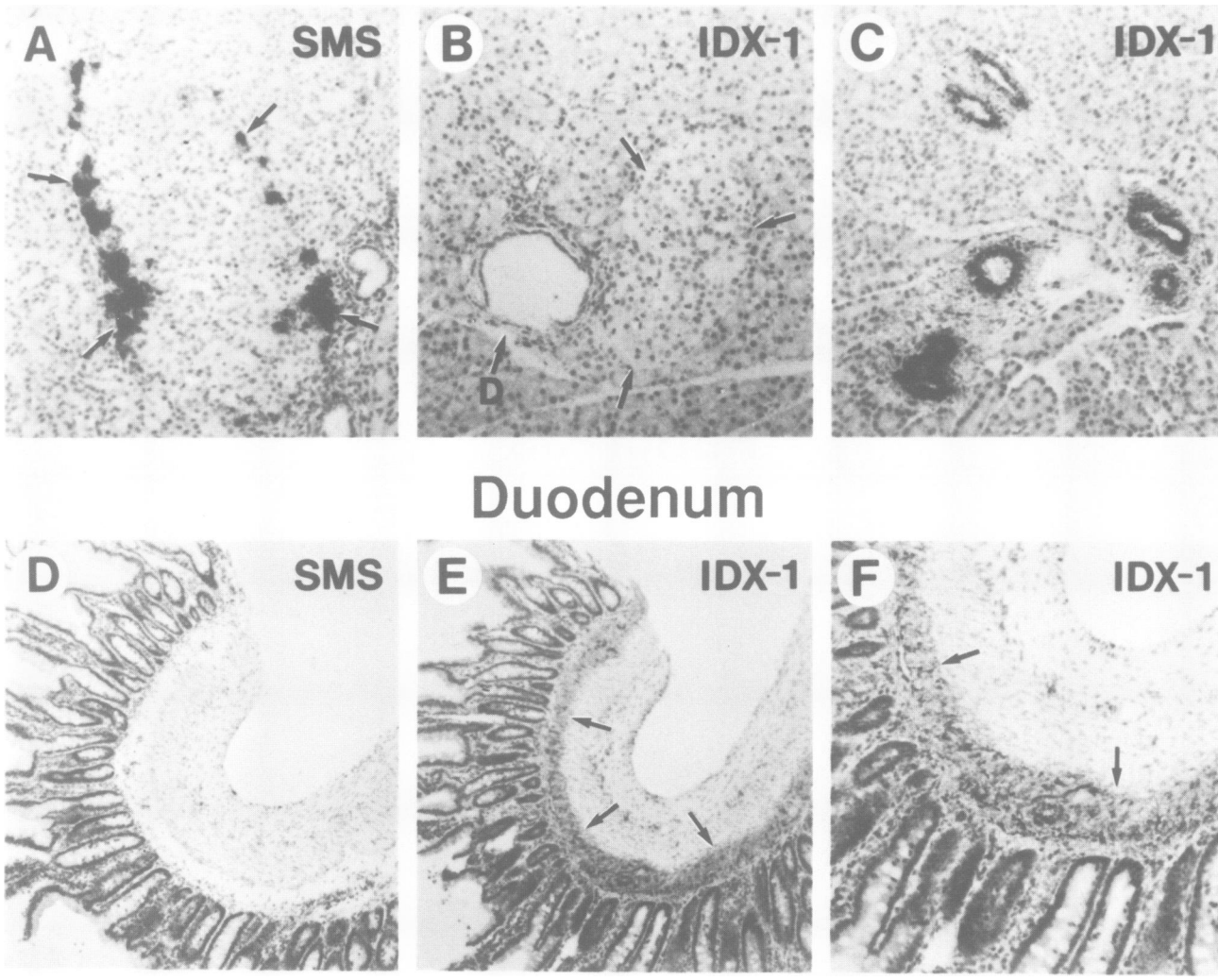


Fig. 5. *In situ* hybridization of somatostatin (SMS, panels A and D) or IDX-1 (B, C, E and F) mRNAs in rat pancreas (A–C) and duodenum (D–F). (D, pancreatic duct; A–C and F, ×1480; D and E, ×740)

(CAT). SMS900CAT contains the three IDX-1 binding sites identified by EMSA (SMS-TAAT1, SMS-TAAT2 and SMS-UE-B; Figure 6B and C), and is transactivated by IDX-1 (4-fold; Figure 7B). IDX-1 also transactivates the reporter plasmid SMS900(Δ−250/−120)CAT which contains the three IDX-1 binding sites, but lacks transcriptional silencer elements (Vallejo *et al.*, in preparation) located between nucleotides −250 and −120 (5-fold, Figure 7B). Additional experiments were conducted to determine to what extent the SMS-TAAT1 and SMS-TAAT2 sites contribute to IDX-1 activation of SMS900CAT. First, the −900 to −250 portion of the rat somatostatin gene (containing SMS-TAAT1 and SMS-TAAT2 but not SMS-UE) was inserted upstream of the heterologous thymidine kinase (TK) promoter in the reporter plasmid pUTKAT (Prost and Moore, 1986). This plasmid, SMS(−900/−250)TKCAT, is transactivated by IDX-1 (4-fold, Figure 7B), whereas the parent plasmid, pUTKAT (TKCAT), is not. These results indicate that the −900 to −250 portion of the rat somatostatin gene can confer IDX-1 responsiveness upon a heterologous promoter, and are consistent with the proposal that SMS-TAAT1 and/or SMS-TAAT2 may mediate IDX-1 transactivation of SMS900CAT. To test this hypothesis further, nucleotide substitution mutations were introduced into the SMS-TAAT1, SMS-TAAT2 and SMS-UE-B sequences within the SMS900CAT reporter plasmid. As shown in Figure 7C, the SMS-TAAT1 sequence appears to be required for full activation of SMS900CAT by IDX-1, whereas the SMS-TAAT2 and SMS-UE-B sites are less critical. Finally, in experiments with reporter plasmids containing multimerized IDX-1 binding sites, IDX-1 potently transactivates a reporter construct containing the multimerized SMS-TAAT1 site (34-fold, Figure 7D) but does not increase transcription of constructs containing multimerized SMS-TAAT2 or INS1-FLAT sequences. Taken together, the results from these cotransfection experiments suggest that IDX-1 transactivates expression of the somatostatin promoter reporter constructs mainly through the upstream SMS-TAAT1 sequence.

RT-PCR analysis of homeodomain mRNAs expressed in rat pancreatic islets

Rat islet cDNA preparations were PCR-amplified with degenerate primers designed to anneal to the most highly

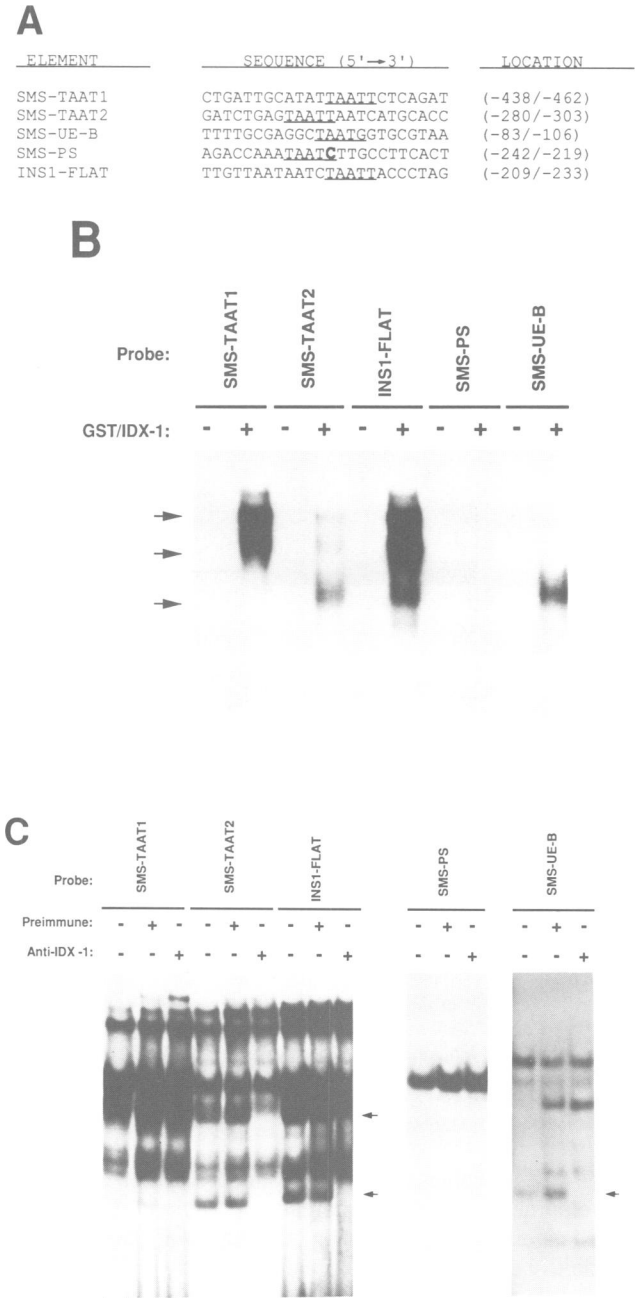


Fig. 6. Sequence-specific DNA binding by IDX-1. (**A**) Sequences and locations (relative to transcription start site) of DNA elements in the rat somatostatin (SMS) gene 5′ flanking region that contain potential IDX-1 binding sites. Also shown is the sequence of the Far-linked AT-rich element (INS1-FLAT), a homeodomain binding sequence in the rat insulin 1 gene mini-enhancer (German *et al.*, 1991), and also a potential IDX-1 binding site. The 5′-TAAT-3′ motifs are underlined. (**B**) Electrophoretic mobility shift assay (EMSA) with bacterially expressed IDX-1 and ^{32}P-labelled oligonucleotide probes corresponding to the sequences shown in Figure 5A. (**C**) EMSA with RIN1027-B2 nuclear extracts and the oligonucleotide probes shown in Figure 5A. DNA−protein binding reactions were performed either under standard conditions, or with the addition of preimmune serum or IDX-1 antiserum. Arrows indicate the DNA−protein complexes that are selectively disrupted by IDX-1 antiserum.

conserved regions of the Antp homeodomain (PCRHD1 and PCRHD2 in Figure 8). PCR products were subcloned into pBluescript(KS) and sequenced. Twelve different homeodomain sequences were found among the 80 subclones that were sequenced. The translated amino acid sequences of the 12 homeodomain cDNAs from rat islets are shown in Figure 8, compared with the corresponding region of the Antp homeodomain with which most share significant homology. Figure 8 also lists the frequencies of occurrence and highest GenBank similarities for these sequences. Sequence # 1 corresponds to the IDX-1 homeodomain and is clearly the most frequently encountered Antp-like homeodomain in rat islet cDNA preparations (45 out of 80 sequences). Ten of the 11 additional sequences are likely to be the rat homologues of previously cloned homeodomain mRNAs. These include Cdx-4 (sequence # 16), Hox1.4 (# 17), CHox-7 (# 22), Hox2.6 (# 24), Cdx-3 (# 27), Cdx-1 (# 35), Hox4.3 (# 39), Hox1.11 (# 43), Hox4A (# 48) and Hox1.3 (# 82). With the exception of Cdx-3 (German *et al.*, 1992b), this is the first evidence for expression of these homeodomain mRNAs in rat islets. Sequence # 19 currently has no significant matches to sequences contained within GenBank, and appears to represent an additional new rat homeodomain mRNA.

Discussion

IDX-1 is a newly discovered homeodomain transcription factor isolated from a somatostatin-producing, rat islet-derived cell line, RIN1027-B2. We show here that IDX-1 is expressed in pancreatic islets and ducts, and in the duodenum, and binds to and activates transcription from regulatory sequence elements in the 5′ flanking region of the rat somatostatin gene. IDX-1 appears to be the first homeodomain protein identified thus far to regulate somatostatin gene transcription. IDX-1 contains an Antp-type homeodomain flanked by proline-rich regions in both the amino and carboxyl domains of the protein. The homeodomain of IDX-1 is presumably involved in sequence-specific DNA binding, whereas, as has been determined for the transcription factor CTF/NF-1 (Mermod *et al.*, 1989), the proline-rich regions in IDX-1 may function in transcriptional activation.

Several IDX-1 binding site elements are identified in the rat somatostatin gene promoter and at least one site is identified in the rat insulin 1 gene 5′ flanking region. Examination of the sequence of the predicted IDX-1 DNA recognition helix (helix 3 of the homeodomain) revealed that there is a single amino acid difference between IDX-1 and Antp (His44 in IDX-1 versus Gln44 in Antp; number refers to position within the homeodomain). Furthermore, Gln44 is highly conserved among all Antp-like homeodomains (Scott *et al.*, 1989). This raised the possibility that IDX-1 may have DNA binding properties different from those of other Antp-like homeoproteins. However, results from our EMSA/PCR-based selection and amplification of IDX-1 binding sites (data not shown) and EMSA with several different potential IDX-1 binding sites indicated that IDX-1 has DNA binding specificity similar to that of other Antp-like homeoproteins (i.e. preference for binding to sites containing a 5′-TAAT(T/G)-3′ sequence; Desplan *et al.*, 1988; Scott *et al.*, 1989; Ekker *et al.*, 1991). Thus, it appears that His44 does not detectably alter the DNA binding sequence specificity of IDX-1, a finding consistent with studies indicating the apparent unimportance of amino acid residue 44 in homeodomain−DNA interactions (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Furukubo-Tokunaga *et al.*, 1992).

The sequence similarity in the homeodomain regions of

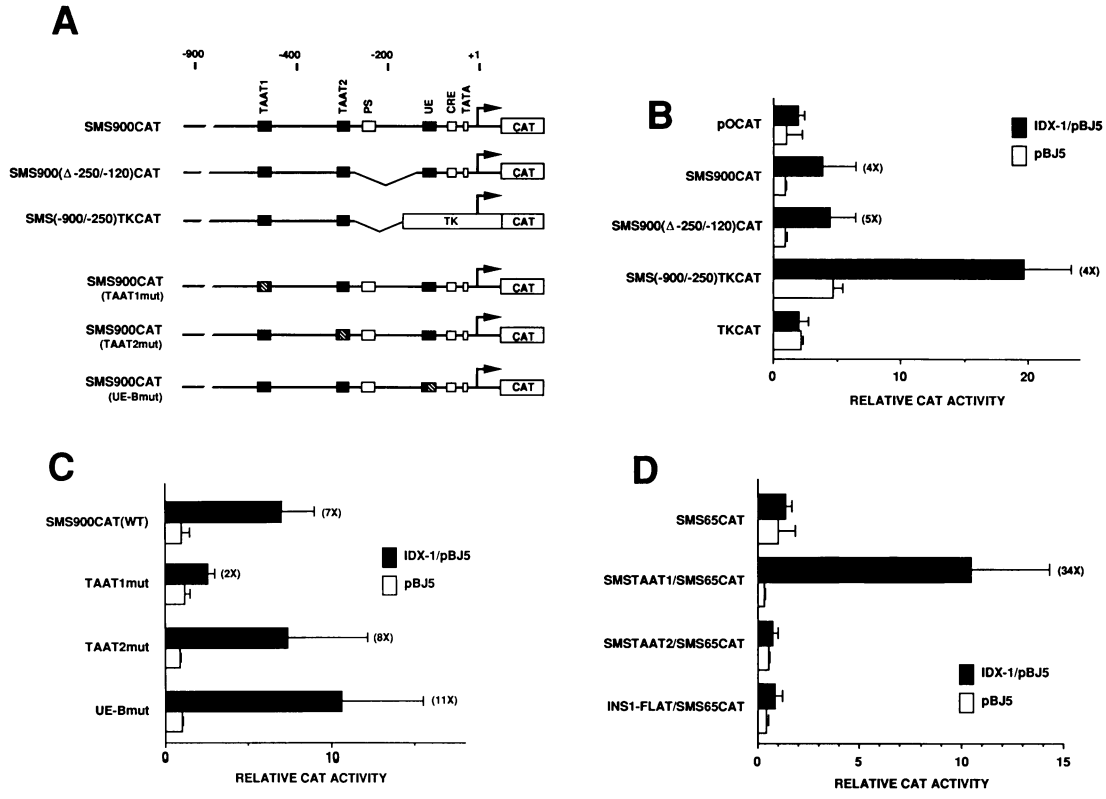


Fig. 7. Transcriptional activation of somatostatin promoter reporter constructs by IDX-1. (**A**) Schematic representation of the somatostatin CAT reporter constructs used in panels B and C. IDX-1 binding sites identified by EMSA (TAAT1, TAAT2 and UE-B) are shown (closed boxes) along with other known regulatory elements (open boxes) in the rat somatostatin promoter. PS, proximal silencer (Vallejo *et al.*, in preparation); CRE, cyclic AMP response element (Powers *et al.*, 1989); +1, transcription start site; TK, thymidine kinase. Clustered point mutations of SMS-TAAT1, SMS-TAAT2 and SMS-UE-B are designated by cross-hatched boxes. (**B**) NIH-3T3 cells were cotransfected with an expression vector encoding IDX-1 (IDX-1/pBJ5) and CAT reporter constructs containing portions of the rat somatostatin 5′ flanking region. Control transfections were performed using the pBJ5 expression vector lacking a cDNA insert. CAT activities for each experiment were normalized relative to the control values obtained for the parent reporter plasmid, pOCAT. Results represent the mean (±SEM) from three or four separate transfections with duplicate plates of cells. Numbers in parentheses indicate fold induction by IDX-1. (**C**) Results from cotransfection experiments in NIH-3T3 cells with IDX-1/pBJ5 and the SMS900CAT reporter plasmids containing wild-type or mutated IDX-1 binding sites shown in panel A. CAT activities for each experiment were normalized relative to control values obtained for SMS900CAT(WT). (**D**) Results from cotransfection experiments with IDX-1/pBJ5 and CAT reporter constructs containing multimerized IDX-1 binding sites inserted into the plasmid SMS65CAT (Powers *et al.*, 1989). CAT activities for each experiment were normalized relative to the control values obtained for SMS65CAT.

IDX-1 and XlHbox8 (Wright *et al.*, 1988) raises the possibility that IDX-1 is the rat homologue of XlHbox8. The fact that only a partial sequence is available for XlHbox8 limits direct sequence comparison. Within the available sequence for the homeodomain region, IDX-1 and XlHbox8 share 100% amino acid and 80% nucleic acid sequence similarity. However, just carboxy-proximal to the homeodomains the sequences of the two proteins diverge completely. Furthermore, although there are strong similarities in the tissue-specific patterns of expression of IDX-1 and XlHbox8, there are also some key differences, particularly with regard to expression within pancreatic islets. XlHbox8 is expressed in a narrow strip of endoderm that gives rise to the duodenum and pancreatic anlaga in *Xenopus* embryos. Expression persists within these structures through maturation of *Xenopus* to adulthood. The highest levels of XlHbox8 in the adult *Xenopus* are in the epithelial cells of the duodenum and the pancreatic secretory duct system. No XlHbox8 was detected in pancreatic islet cells (Wright *et al.*, 1988). Although the developmental expression of IDX-1 has not been characterized, we observe that IDX-1 is detected in fetal intestine and adult duodenum, and is readily detected in pancreatic islets and ducts and mesenchymal cells surrounding the pancreatic ducts. It may be that XLHbox8 and IDX-1 serve similar functions during the early development of the pancreas in *Xenopus* and rat, respectively, but differ during later stages of development. Definitive establishment of a relationship between IDX-1 and XlHbox8 will require complete sequence information for XlHbox8, and a detailed study of IDX-1 expression during embryonic development of the rat.

The observation that IDX-1 expression is detected only in pancreatic islets and ducts and in the duodenum is intriguing, as it is well established that the pancreas is derived from the duodenum during embryonic development (Pictet and Rutter, 1972), and it is believed that islet cells arise from stem cell populations in or around the pancreatic ducts (Yoshinari and Diakoku, 1982; Teitelman and Lee, 1987; Alpert *et al.*, 1988; Dudek *et al.*, 1991). This pattern of expression suggests a possible role for IDX-1 in development of the endocrine pancreas. A number of polypeptide hormone genes are expressed both in islets and in specialized cells within the small intestine of rodents. These genes include insulin, somatostatin, glucagon and pancreatic polypeptide (Green *et al.*, 1989). Furthermore, islet cell tumours and transformed islet cell lines often express genes encoding polypeptide hormones normally produced in the gut (Philippe *et al.*, 1987). These observations, coupled now with the

```
                    20        30        40        50
                     |         |         |         |
            Antp  ELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQN

Seq#   Freq.                                          Closest Genbank Matches (aa, na); ref.

1      (45/80)   ------L--K-IS-P--V-L-VM-N----H-------  IDX-1; XlHbox8 (100%, 80%); Wright et al., 1989
16     (6/80)    -------C-R-I-I--KS-L-VN-G-S---V------  Mouse Cdx-4 (100%, 98%); Gamer and Wright, 1993
17     (2/80)    ----------------------T---S---V------  Mouse Hox1.4 (100%, 95%); Duggal et al., 1987
19     (3/80)    ------ETQK--SPPE-KRL-KM-Q-S---V------  No Significant Matches
22     (1/80)    -------CKK--SLTE-SQ-----K-S-V-V------  CHox7 (100%, 98%); Fainsod and Greunbaum, 1989
24     (6/80)    -------Y---------V--------S----------  Mouse Hox2.6 (100%, 97%); Graham et al., 1988
27     (4/80)    -------YS--I-I--KA-L-AT-G-S---V------  Hamster Cdx-3 (100%, 92%); German et al., 1992
35     (5/80)    -------YS--I-I--KS-L-AN-G-----V------  Mouse Cdx-1 (100%, 92%); Duprey et al., 1988
39     (3/80)    ------L--P----K----VS-T-G-----V------  Mouse Hox4.3 (100%, 99%); Nazarali et al., 1992
43     (2/80)    ---------K--C-P--V---AL-D-----V------  Rat Hox1.11 (100%, 100%); Patel et al., 1992
48     (1/80)    ------------C-P--V-M-NL-N------------  Mouse Hox-4A (100%, 96%); Lonai et al., 1987
82     (1/80)    --------------------------S----------  Mouse Hox1.3 (100%, 97%); Odenwald et al., 1987

                 ----->                          <-----
                 PCRHD1                          PCRHD2
```

Fig. 8. Translated amino acid sequences of Antp-like homeodomain mRNAs PCR-amplified from rat islet cDNA. Dashes denote sequence identity with the Antp homeodomain shown at the top. Locations of PCRHD1 and PCRHD2 amplimers are shown below. Numbers above Antp sequence indicate amino acid position within the homeodomain. Also shown are the frequencies of occurrence and closest sequence matches found in GenBank. Frequency of occurrence refers to the number of times a given homeodomain sequence was obtained from sequencing 80 separate DNA minipreps. Numbers in parentheses indicate percent sequence similarity between the sequences amplified from rat islet cDNA and the GenBank entries for previously cloned homeodomains (amino acid similarity; nucleic acid similarity). Sequence #1, which represents >50% of the cloned sequences (45/80), is IDX-1. Sequence comparisons were confined to the regions between amino acids 21 and 46 of the homeodomain. GenBank searches and sequence comparisons used BLAST software.

shared expression of selected homeodomain transcription factors such as IDX-1, are consistent with the longstanding notion that the endocrine pancreas represents a specialization of the secretory capacity of the intestinal epithelium (Barrington, 1964; Falkmer and Patent, 1972).

Several other homeodomain mRNAs have been detected in rat or mouse intestine. These include Cdx-1, -2, -3 and -4, and Hox1.3, 1.6, 1.7, 2.3, 2.5, 3.1 and 3.2 (Baron *et al.*, 1987; James and Kazenwadel, 1991; German *et al.*, 1992; Gamer and Wright, 1993). Of the genes encoding these homeodomain mRNAs, only Cdx-1, Cdx-2 and Hox1.3 are known to be expressed in the duodenum, and these are expressed at higher levels in the colon than the duodenum (James and Kazenwadel, 1991). Most of the homeodomain genes expressed in intestine have been detected in multiple segments of the small and/or large intestine. Within the gastrointestinal tract, IDX-1 differs from these other homeodomain mRNAs in that its expression appears to be restricted exclusively to the duodenum (Figure 2B). IDX-1 may thus play a role in development and differentiation of the proximal small intestine and specialized structures that arise from it, such as the pancreatic islets.

At least 11 other Antp-like homeodomain mRNAs are expressed in pancreatic islets (Figure 8). PCR amplification of homeodomain sequences for Hox1.3, Hox1.4, Hox1.11, Hox2.6, Hox4A, Hox4.3, CHox-7, Cdx-1, Cdx-3 and Cdx-4 from rat islet cDNA indicates that mRNAs for these previously characterized genes are expressed in rat islets. Of these, Cdx-3 is the only Antp-like homeoprotein that has previously been shown to be expressed in rat islets. Cdx-3 has been shown to activate weakly reporter constructs containing the FLAT element of the rat insulin 1 gene (German *et al.*, 1992b), and may therefore play a role in islet hormone gene regulation. One of the other genes, Hox1.11, has recently been shown to be expressed in vascular smooth muscle (Patel *et al.*, 1992). It is possible that its presence in rat islet cDNA may be due to the small amount of vascular smooth muscle contained in islet arterioles. If so, Hox1.11 may be involved in the development of islet vasculature. For the nine additional Antp-like homeodomain mRNA sequences in rat islets, it remains to be shown whether the corresponding homeodomain proteins are produced and are involved in islet gene expression or development and differentiation.

IDX-1 activates transcription of the rat somatostatin gene. In cotransfection experiments in NIH-3T3 cells, this activation appears to occur mainly through the SMS-TAAT1 element located 450 nucleotides upstream of the transcription start site. Three lines of evidence support this proposal. First, a fragment of the rat somatostatin promoter containing the SMS-TAAT1 and SMS-TAAT2 sequences renders the heterologous TK promoter responsive to transactivation by IDX-1. Second, mutation of SMS-TAAT1 sequence within the SMS900CAT reporter attenuates activation by IDX-1, whereas mutations of the SMS-TAAT2 or SMS-UE-B sites appear not to alter activation. Third, multimerized SMS-TAAT1 sites, but not SMS-TAAT2 sites, are activated by IDX-1. That the activation of SMS-TAAT1 is greater than that of SMS-TAAT2 may be related to differences in IDX-1 binding affinities. In EMSA, bacterially expressed IDX-1 appears to bind SMS-TAAT1 better than SMS-TAAT2 (Figure 6B). However, IDX-1 also binds well to the INS1-FLAT probe, yet fails to activate a reporter construct containing multimerized INS1-FLAT sites (Figure 7D).

Consequently, differences in binding alone may not account for differences in transactivation of these related sequences. The INS1-FLAT site has been shown to be composed of adjacent positive and negative domains such that under certain conditions transcriptional activation requires mutation of the suppressor portion of the element (German *et al.*, 1992a). A similar situation could exist for the SMS-TAAT2 sequence, so that the failure of IDX-1 to transactivate through this element under the conditions used in these experiments should not be interpreted to mean that this site cannot function in transcriptional regulation of the somatostatin gene under other circumstances.

IDX-1 shares several features with three other DNA binding proteins previously characterized in islet-derived cell lines. IDX-1 mRNA and protein are found in RIN1027-B2 and RIN1046-38 cells but not in InR1G9 cells (Figure 3). This suggests that IDX-1 is expressed in β- and δ-cells but not in α-cells. A similar pattern of expression was described for IPF-1, a protein expressed in β-TC1 but not α-TC1 cells that binds to the sequence 5′-CC(C/T)TAATGGG-3′ of the rat insulin I gene promoter (Ohlsson *et al.*, 1991). IDX-1 may also be related to IUF-1, a factor identified in RINm5F and HIT-T15 cells that binds to the sequence 5′-C(T/C)CTA-ATG-3′ in the human insulin gene enhancer (Boam and Docherty, 1989). In addition to sharing DNA binding site sequences and being expressed in similar islet cell phenotypes, IDX-1, IPF-1 and IUF-1 have comparably fast electrophoretic mobilities in EMSA, suggesting that they may have similar molecular weights. IDX-1 also shares DNA binding sequence specificity, electrophoretic mobility in EMSA and cell-type expression with a factor that binds to the B domain of the SMS-UE, a cell-specific enhancer in the rat somatostatin gene 5′ flanking region (Vallejo *et al.*, 1992a,b).

Based upon its relative abundance, pattern of expression and transcriptional regulatory activity, it seems reasonable to postulate that IDX-1 may play a role in development and differentiation of the proximal small intestine and the endocrine pancreas, and that IDX-1 may be an important transcriptional regulator in differentiated islet cells.

Materials and methods

cDNA cloning

A λgt11 cDNA library from RIN1027-B2 cells was screened at low stringency with a ^{32}P-labelled single-stranded degenerate oligonucleotide (HB-1) corresponding to the most highly conserved amino acid residues in helix 3 of the homeodomain (Bürglin *et al.*, 1989). cDNA inserts from plaque-purified phage were amplified by PCR, subcloned into pBluescript(KS) (Stratagene, Torrey Pines, CA) and sequenced by the dideoxy chain termination method (Sequenase, USB, Cleveland, OH) using T3 and T7 primers. Initial screening with ^{32}P-labelled HB-1 yielded 18 plaque-purified recombinant phage. Two of these contained identical 0.8 kb cDNA inserts with high sequence similarity to XlHbox8, an endoderm-specific homeobox transcription factor expressed in the pancreatic anlagen of *X.laevis* (Wright *et al.*, 1987). These cDNA inserts correspond to partial IDX-1 clones. One of these cDNAs was labelled with ^{32}P and used to rescreen the RIN1027-B2 cDNA library to isolate the 1.4 kb IDX-1 cDNA.

Cell culture

RIN1027-B2, RIN1046-38 (Chick *et al.*, 1977; Philippe *et al.*, 1987), InR1G9 (Takaki *et al.*, 1986), HepG2, Hep3B, JEG-3 and HeLa cells (ATCC HB8065, HB8064, HTB36 and CCL2, respectively) were cultured at 37°C in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 100 U of penicillin per ml, and 100 μg of streptomycin per ml. NIH-3T3 cells (ATCC CRL1658) were cultured in DMEM with the same antibiotics and 10% calf serum. βTC1 cells (Efrat *et al.*, 1988) were cultured in RPMI-1640 with 10% fetal bovine serum and antibiotics. All cell culture reagents were obtained from Gibco (Grand Island, NY).

Northern blots

Total RNA prepared from rat tissues or cell lines (Ausubel *et al.*, 1992) was fractionated on formaldehyde−1% agarose gels, then transferred to nylon membranes (MSI, Westborough, MA). RNA bound to the membranes was hybridized overnight at 48°C with random primed ^{32}P-labelled cDNA probes for rat IDX-1 or chicken β-actin (Cleveland *et al.*, 1980). Membranes were washed and dried, then analysed by autoradiography.

Western immunoblots

Antiserum to IDX-1 was raised in rabbits immunized (Hazelton Research Products, Denver, PA) with a bacterially expressed recombinant glutathione S-transferase (GST)−IDX-1 fusion protein (Smith and Johnson, 1988) containing amino acid residues 164−283 of rat IDX-1. The R1090 CREB antiserum has been described (Lee *et al.*, 1990). Crude nuclear lysates were prepared from rat duodenum and jejunum by isolating nuclei from these tissues through a sucrose cushion (Gorski *et al.*, 1986), then lysing the nuclei in SDS sample buffer (Laemmli, 1970). Pancreatic islet whole cell lysates were prepared by lysing freshly isolated rat islets in SDS sample buffer. Whole cell lysates from RIN1027-B2 cells were prepared in a similar manner. Lysates were sonicated then cleared by centrifugation (10 000 *g*, 4°C). Nuclear extracts were prepared from cultured cells by the method of Dignam *et al.* (1983). Extracts and lysates were fractionated on SDS−polyacrylamide gels and electroblotted onto nitrocellulose membranes (MSI). The membranes were incubated with primary antisera (1:20 000) and subsequently with an alkaline phosphatase-conjugated goat anti-rabbit secondary antibody (Bio-Rad Laboratories, Richmond, CA). Immunoreactive proteins were visualized using the ECL chemiluminescent detection system (Amersham Inc., Arlington Heights, IL).

Immunohistochemistry

RIN1027-B2 cells grown on chamber slides (Falcon, Oxnard, CA), or freshly cut frozen pancreas and duodenum sections, were fixed with 4% paraformaldehyde in phosphate-buffered saline, permeabilized in acetone at −20°C and incubated overnight at 4°C with IDX-1 antiserum. To visualize IDX-1 protein, slides were incubated with a biotinylated secondary antibody and then with an avidin-biotinylated horseradish peroxidase complex (Vectastain ABC System, Vector Laboratories, Burlingame, CA).

In situ hybridization

Rat pancreas and duodenum samples were excised and fixed overnight at 4°C with 4% paraformaldehyde in phosphate-buffered saline. Tissue samples were dehydrated and embedded in paraffin. Sections were cut and *in situ* hybridizations performed with ^{35}S-labelled cDNA or cRNA probes for rat somatostatin or IDX-1 mRNA (Simmons *et al.*, 1989).

RT-PCR amplification of homeobox cDNAs from rat pancreatic islets

Rat pancreatic islets were isolated by collagenase digestion and manual selection/transfer (Lacy and Kostianovsky, 1967). Total RNA was prepared from ~2000 islets (Ausubel *et al.*, 1992). First strand cDNA synthesis was performed as described by Ausubel *et al.* (1992) using 5 μg islet total RNA and AMV reverse transcriptase. One microlitre of this cDNA preparation was used as a template for PCR amplification. Amplimer sequences, PCR conditions and subcloning procedures were exactly as described by James and Kazenwadel (1991). The PCR amplimers contain *Bam*HI or *Eco*RI restriction enzyme sites and generate PCR products of 120 bp in length from homeodomain cDNA templates. The PCR products represent mixed populations of homeodomain sequences, which were digested with *Bam*HI and *Eco*RI, ligated into *Bam*HI/*Eco*RI-cut pBluescript(KS) and used to transform JM109 bacteria. Individual bacterial colonies were picked, then miniprep plasmid DNAs were prepared and sequenced. Sequences were compared with GenBank entries using the BLAST network service through the National Center for Biotechnology Information (NCBI).

Transfections and CAT assays

A eukaryotic expression vector for IDX-1 was constructed by inserting the 1.4 kb IDX-1 cDNA into the plasmid pBJ5 (Lin *et al.*, 1990). Several different reporter constructs were used in cotransfection experiments with IDX-1/pBJ5 expression vector. The SMS900CAT plasmid consists of nucleotides −900 to +54, relative to the transcription start site, of the rat somatostatin gene linked 5′ to the coding sequence of the bacterial CAT gene (Powers *et al.*, 1989). The plasmid SMS900(Δ−250/−120)CAT is

identical to SMS900CAT with the exception of an internal deletion between nucleotides −250 and −120. This deletion removes a putative transcriptional suppressor region located between nucleotides −237 and −189 of the rat somatostatin gene (Vallejo *et al.*, in preparation). SMS(−900/−250)TKCAT contains nucleotides −900 to −250 of the rat somatostatin gene linked 5′ to the herpes simplex thymidine kinase promoter in the CAT reporter plasmid pUTKAT (Prost and Moore, 1986). The plasmids SMS900CAT(TA-AT1mut), SMS900CAT(TAAT2mut) and SMS900CAT(UE-Bmut) were created using oligonucleotide-directed mutagenesis (Kunkel, 1985) to introduce clustered nucleotide substitutions in the IDX-1 binding sites located at −450, −290 and −90, respectively, within the SMS900CAT reporter. For SMS-TAAT1, the central 5′-ATTAAT-3′ was converted to 5′-GCC-GGC-3′. Likewise, SMS-TAAT2 and SMS-UE-B were changed from 5′-TAATTA-3′ and 5′-CTAATG-3′ to 5′-CGGCCG-3′ and 5′-TCG-GCA-3′, respectively. These mutations abolish IDX-1 binding in EMSA (not shown). Sequences of the mutations were confirmed by dideoxy sequencing. Reporter plasmids were also constructed that contained multimerized double-stranded oligonucleotides corresponding to IDX-1 binding sites (see below for oligonucleotide sequences) inserted into the *Bam*HI sites of the plasmids SMS65CAT (Powers *et al.*, 1989). Sequence, copy number and orientation of the oligonucleotides were confirmed by restriction digestion and dideoxy sequencing. The SMSTAAT1/SMS65CAT construct contains six copies of the SMS-TAAT1 oligonucleotide, linked head-to-tail. The SMS-TAAT2/SMS65CAT and INS1-FLAT/SMS65CAT constructs contain three and five head-to-tail copies of their respective oligonucleotides.

NIH-3T3 cells were transfected using $CaPO_4$ (Ausubel *et al.*, 1992) with 5 μg of reporter constructs and 5 μg of IDX-1/pBJ5 expression plasmid. Control transfections were performed using pBJ5 with no cDNA insert. Forty-eight hours after transfections, cell extracts were prepared and CAT activities in aliquots of the extracts were determined (Ausubel *et al.*, 1992).

DNA−protein binding assays

Electrophoretic mobility shift assays (EMSA) were performed as described by Ron *et al.* (1990) using bacterially expressed IDX-1 or nuclear extracts prepared from RIN1027-B2 cells (Dignam *et al.*, 1983). PCR was used to generate a bacterial expression vector encoding the full-length IDX-1 protein fused to GST. An oligonucleotide amplimer was designed to create an *Nco*I site at the initiation methionine of the IDX-1 cDNA. This primer was used in conjunction with the T3 primer to amplify the IDX-1 cDNA insert from an IDX-1/pBS(KS) subclone. The PCR product was digested with *Nco*I and *Xho*I, gel-purified and ligated into *Nco*I/*Xho*I-cut pGEX-KG (Pharmacia Biotech Inc., Piscataway, NJ). Expression and purification of the GST/IDX-1 fusion protein was performed as described by Smith and Johnson (1988).

Nuclear extracts from RIN1027-B2 cells or purified, bacterially expressed GST/IDX-1 were incubated with the following ^{32}P-labelled double-stranded oligonucleotide probes: 1, INS1-FLAT, the Far-linked AT-rich element of the rat insulin 1 gene (German *et al.*, 1992a) (5′-GATCCTTGTTAATAA-TCTAATTACCCTAGGTCTAA-3′); 2, SMS-TAAT1, a FLAT-like element (5′-GATCCCTGATTGCATATTAATTCTCAGATA-3′) located at nucleotides −438 to −461, relative to the transcription start site, of the rat somatostatin gene; 3, SMS-TAAT2, a second FLAT-like element in the rat somatostatin gene (nucleotides −290 to −303; 5′-GATCCGATC-TCAGTAATTAATCATGCACCA-3′); 4, SMS-UE-B, the B domain of the rat somatostatin upstream enhancer element (Vallejo *et al.*, 1992a,b) (nucleotides −83 to −106; 5′-GATCCGCGAGGCTAATGGTGCGTA-AAAGCACTGGTGA-3′); 5, SMS-PS, a transcriptional silencer element in the rat somatostatin gene (M.Vallejo, C.P.Miller, W.Beckman and J.K.Habener, in preparation; nucleotides −219 to −233; 5′-GATCCA-GGCAAGATTATTTGGTCA-3′). These oligonucleotides all contain 5′-TAAT-3′ motifs that are typical of DNA sites recognized by homeodomain transcription factors (Desplan *et al.*, 1988; Scott *et al.*, 1989; Ekker *et al.*, 1991). After incubation with the probes, DNA−protein complexes were resolved on non-denaturing polyacrylamide gels, then visualized by autoradiography of the dried gels.

Acknowledgements

We are grateful to J.A.Chafitz, J.C.Lin and M.Lesieur for expert technical support, Drs W.Walker, M.Vallejo and C.Girardet for helpful discussions, and H.Hermann for special assistance. We also thank Drs A.Joyner (University of Toronto) and W.A.Miller (SUNY, Buffalo) for advice and help with *in situ* hybridization procedures, and Dr Mark Davis (Stanford University) for providing the pBJ5 expression vector. These studies were supported in part by USPHS grants DK30457 and DK30834. J.F.H. is an investigator of the Howard Hughes Medical Institute.

References

Alpert,S., Hanahan,D. and Teitelman,G. (1988) *Cell*, **53**, 295−308.
Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1992) *Short Protocols in Molecular Biology*. 2nd edn. John Wiley and Sons, New York, NY.
Barrington,E.J.W. (1964) *Hormones and Evolution*. English University Press, London.
Bellmann,R. and Werr,W. (1992) *EMBO J.*, **11**, 3367−3374.
Blumberg,B., Wright,C.V.E., DeRobertis,E.M. and Cho,K.W.Y. (1991) *Science*, **253**, 194−196.
Boam,D.S.W. and Docherty,K. (1989) *Biochem. J.*, **264**, 233−239.
Baron,A., Featherstone,M.S., Hill,R.E., Hall,A., Galliot,B. and Duboule,D. (1987) *EMBO J.*, **6**, 2977−2986.
Bürglin,T.R., Finney,M., Coulson,A. and Ruvkun,G. (1989) *Nature*, **341**, 239−243.
Chick,W.L., Warren,S., Chute,R.N., Like,A.A., Lauris,V. and Kitchen,K.C. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 628−632.
Cleveland,D.W., Lopata,M.A., MacDonald,R.J., Cowan,N.J., Rutter,W.J. and Kirschner,M.W. (1980) *Cell*, **20**, 95−105.
Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081−1090.
Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.
Dudek,R.W., Lawrence,I.E.,Jr, Hill,R.S. and Johnson,R.C. (1991) *Diabetes*, **40**, 1041−1048.
Efrat,S., Linde,S., Kofod,H., Spector,D., Delannov,M., Grant,S., Hanahan,D. and Baekkeskov,S. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 9037−9041.
Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10**, 1179−1186.
Emens,L.A., Landers,D.W. and Moss,L.G. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 7300−7304.
Falkmer,S. and Patent,G.J. (1972) In Steiner,D.F. and Freinkel,N. (eds), *Handbook of Physiology*. American Physiological Society, Washington, DC, Section 7, Vol. 1, pp. 1−23.
Furukubo-Tokunaga,K., Muller,M., Affolter,M., Pick,L., Kloter,U. and Gehring,W.J. (1992) *Genes Dev.*, **6**, 1082−1096.
Gamer,L.W. and Wright,C.V.E. (1993) *Mech. Dev.*, **43**, 71−81.
Gehring,W.J. (1987) *Science*, **236**, 1245−1252.
German,M.S., Moss,L.G., Wang,J. and Rutter,W.J. (1992a) *Mol. Cell. Biol.*, **12**, 1777−1788.
German,M.S., Wang,J., Chadwick,R.B. and Rutter,W.J. (1992b) *Genes Dev.*, **6**, 2165−2176.
Gorski,K., Carneiro,M. and Schibler,U. (1986) *Cell*, **47**, 767−776.
Green,D.W., Gomez,G. and Greeley,G.H. (1989) *Gastroenterol. Clin. North Am.*, **18**, 695−733.
Han,J.H., Rall,L. and Rutter,W.J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 110−114.
Holland,P.W.H. and Hogan,B.L.M. (1988) *Genes Dev.*, **2**, 773−782.
James,R. and Kazenwadel,J. (1991) *J. Biol. Chem.*, **266**, 3246−3251.
Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579−590.
Kunkel,T.A. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 488−492.
Lacy,P.E. and Kostianovsky,M. (1967) *Diabetes*, **16**, 35−39.
Laemmli,U.K. (1970) *Nature*, **277**, 680.
Lee,C.Q., Yun,Y., Hoeffler,J.P. and Habener,J.F. (1990) *EMBO J.*, **9**, 4455−4465.
Levine,E.M. and Schechter,N. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 2729−2733.
Lin,A., Devaux,B., Green,A., Sagerstrom,C., Elliot,J. and Davis,M. (1990) *Science*, **249**, 677−679.
Mermod,N., O'Neill,E.A., Kelly,T.J. and Tjian,R. (1989) *Cell*, **58**, 741−753.
Murtha,M.T., Leckman,J.F. and Ruddle,F.H. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 10711−10715.
Nazarali,A., Kim,Y. and Nirenberg,M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2883−2887.
Nohno,T., Noji,S., Koyama,E., Ohyama,K., Myokai,F., Kuroiwa,A., Saito,T. and Taniguchi,S. (1991) *Cell*, **64**, 1197−1205.
Ohlsson,H., Thor,S. and Edlund,T. (1991) *Mol. Endocrinol.*, **5**, 897−904.
Orci,L. (1982) *Diabetes*, **31**, 538−565.
Otting,G., Qian,Y.Q., Billeter,M., Muller,M., Affolter,M., Gehring,W.J. and Wuthrich,K. (1990) *EMBO J.*, **9**, 3085−3092.
Patel,C.V., Gorski,D.H., LePage,D.F., Lincecum,J. and Walsh,K. (1992) *J. Biol. Chem.*, **267**, 26085−26090.
Philippe,J., Chick,W.L. and Habener,J.F. (1987) *J. Clin. Invest.*, **79**, 351−358.

Pictet,R. and Rutter,W.J. (1972) In Steiner,D.F. and Freinkel,N. (eds), *Handbook of Physiology*. American Physiology Society, Washington, DC, Section 7, Vol. 1, pp. 25–66.
Powers,A.C., Tedischi,F., Wright,K.E., Chan,J.S. and Habener,J.F. (1989) *J. Biol. Chem.*, **264**, 10048–10056.
Prost,E. and Moore,D.D. (1986) *Gene* (Amsterdam), **45**, 107–111.
Renucci,A., Zappavigna,V., Zàkàny,J., Izpisúa-Belmonte,J.-C., Bürki,K. and Duboule,D. (1992) *EMBO J.*, **11**, 1459–1468.
Ron,D., Brasier,A.R., Wright,K.A., Tate,J.E. and Habener,J.F. (1990) *Mol. Cell. Biol.*, **10**, 1023–1032.
Schummer,M., Scheurlen,I., Schaller,C. and Galliot,B. (1992) *EMBO J.*, **11**, 1815–1823.
Scott,M.P., Tamkun,J.W. and Hartzell,G.W.,III (1989) *Biochim. Biophys. Acta*, **989**, 25–48.
Simmons,D.M., Arriza,J.L. and Swanson,L.W. (1989) *J. Histotechnol.*, **12**, 169–181.
Singh,G., Kaur,S., Stock,J.L., Jenkins,N.A., Gilbert,D.J., Copeland,N.G. and Potter,S.S. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 10706–10710.
Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.
Sun,X.-H. and Baltimore,D. (1991) *Cell*, **64**, 459–470.
Takaki,R., Ono,J., Nakamura,M., Yokogawa,Y., Kumae,S., Hiraoka,T., Yamaguchi,K. and Uchida,S. (1986) *In Vitro Cell. Dev. Biol.*, **22**, 120–126.
Teitelman,G. and Lee,J.K. (1987) *Dev. Biol.*, **121**, 454–466.
Vallejo,M., Miller,C.P. and Habener,J.F. (1992a) *J. Biol. Chem.*, **267**, 12868–12875.
Vallejo,M., Penchuk,L. and Habener,J.F. (1992b) *J. Biol. Chem.*, **267**, 12876–12884.
Weeden,C.J., Kostriken,R.G., Matsumura,I. and Weisblat,D.A. (1990) *Nucleic Acids Res.*, **18**, 1908.
Wessells,N.K. and Evans,J. (1968) *Dev. Biol.*, **17**, 413–446.
Wright,C.V., Schnegelsberg,P. and DeRobertis,E.M. (1988) *Development*, **104**, 787–794.
Yoshinari,M. and Diakoku,S. (1982) *Anat. Embryol.*, **165**, 63–70.

Received on August 12, 1993; revised on November 15, 1993

Noted added in proof

The nucleotide sequence data reported in this paper will appear in the EMBL/GenBank databases under the accession number UO4833. During revision of this manuscript cDNAs encoding proteins similar or identical to IDX-1 were reported: STF-1 [somatostatin transactivating factor-1; Leonard *et al.* (1993) *Mol. Endocrinol.*, **7**, 1275–1283] and IPF-1 [insulin prromoter factor-1; Ohlsson *et al.* (1993) *EMBO J.*, **12**, 4251–4259].