MOLECULAR AND CELLULAR BIOLOGY, Oct. 1994, p. 6704–6714
0270-7306/94/$04.00+0

Vol. 14, No. 10

Isolation and Characterization of a Novel Transcription Factor That Binds to and Activates Insulin Control Element-Mediated Expression

GARY L. W. G. ROBINSON,[1] SUSAN R. CORDLE,[1] EVA HENDERSON,[1] P. ANTHONY WEIL,[1] GLADYS TEITELMAN,[2] AND ROLAND STEIN[1]*

Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, Tennessee 37232,[1] and Department of Anatomy and Cell Biology, SUNY Health Sciences Center at Brooklyn, Brooklyn, New York 11203[2]

Received 7 February 1994/Returned for modification 7 April 1994/Accepted 20 July 1994

Pancreatic β-cell-type-specific transcription of the insulin gene is principally regulated by a single *cis*-acting DNA sequence element, termed the insulin control element (ICE), which is found within the 5′-flanking region of the gene. The ICE activator is a heteromeric complex composed of an islet α/β-cell-specific factor associated with the ubiquitously distributed E2A-encoded proteins (E12, E47, and E2-5). We describe the isolation and characterization of a cDNA for a protein present in α and β cells, termed INSAF for insulin activator factor, which binds to and activates ICE-mediated expression. INSAF was isolated from a human insulinoma cDNA library. Transfection experiments demonstrated that INSAF activates ICE expression in insulin-expressing cells but not in non-insulin-expressing cells. Cotransfection experiments showed that activation by INSAF was inhibited by Id, a negative regulator of basic helix-loop-helix (bHLH) protein function. INSAF was also shown to associate in vitro with the bHLH protein E12. In addition, affinity-purified INSAF antiserum abolished the formation of the activator-specific ICE-binding complex. Immunohistochemical studies indicate that INSAF is restricted in terms of its expression pattern, in that INSAF appears to be detected only within the nuclei of islet pancreatic α and β cells. All of these data are consistent with the proposal that INSAF is either part of the ICE activator or is antigenically related to the specific activator required for insulin gene transcription.

The insulin gene is expressed exclusively in the β cells of the islet of Langerhans. This gene accounts for a large fraction of the total gene expression in these cells and has provided an excellent system for studies of cell-type-specific expression. Hanahan in experiments conducted with transgenic mice harboring insulin-simian virus 40 tumor antigen hybrid genes showed that β-cell-specific transcription was regulated by 5′-flanking insulin DNA sequences (21). These studies, which were conducted with the rat insulin II gene, demonstrated that residues which lie between bp −695 and +1 relative to the site of initiation are capable of directing β-cell-type-specific expression. Rodents have two nonallelic insulin genes (I and II), which differ in their number of introns and chromosomal locations (45). The equivalent region of the human insulin gene has also been shown to direct cell-specific expression within transgenic mice (9, 17, 39).

Experiments conducted with insulin- and non-insulin-producing cell lines have demonstrated that the 5′-flanking sequences of the insulin gene that are essential for pancreatic β-cell-type-specific expression reside between bp −340 and −91 relative to the transcription start site (for a review, see reference 44). This region exhibits enhancer-like properties and is regulated by both positively and negatively acting transcription factors (12, 13, 15, 31, 55). Detailed characterization of the insulin gene enhancer indicates that transcription mediated by these sequences is predominantly regulated by a single element (13, 25), whose core motif, 5′-GCCATCTG-3′, is found within the transcription unit of all characterized insulin genes (45). This element, which we refer to as the insulin control element (ICE), is a site of both positive and negative transcriptional control (24, 26, 55). The ICE is also important in homeostatic control of the insulin gene in glucose-treated β cells (19, 40).

The ICE activator appears to be restricted to islet α and β cells (32, 38). Experiments conducted with antisera to the expressed E2A gene products, E12, E47, or ITF-1 (or antigenically related proteins), have demonstrated that these ubiquitously distributed proteins are present in the ICE activator complex (11, 18, 42). (ITF-1 is a generally distributed E47-like factor [23].) Mutagenesis studies have shown that the nucleotides (shown underlined) that define the binding motif for proteins in the basic helix-loop-helix (bHLH) family, CANNTG, are essential for *trans* activation (54). In addition, ICE activation is inhibited in β cells by a negative regulator of bHLH-mediated activation, Id (11). This factor appears to suppress bHLH activation by sequestering the E2A gene products into transcriptionally nonfunctional complexes (5). These results are all consistent with the proposal that the ICE activator is composed, at least in part, of E2A proteins. The restricted distribution of this ICE activator binding activity suggests that an islet α- and β-cell-specific factor may be a partner of the generally expressed E2A proteins in this complex.

In this study, we report the isolation of an ICE binding factor by oligonucleotide screening of a human pancreatic insulinoma λgt11 cDNA expression library. We demonstrate that (over) expression of this gene in transfection experiments stimulates, in a β-cell-specific manner, ICE-dependent transcription and that the activation mediated by this factor can be attenuated by Id. We also find that this protein forms a heteromeric complex with E12 and that affinity-purified antisera to this clone specifically abolish formation of the ICE-binding complex in a mobility shift assay. Immunohistochemical studies demon-

* Corresponding author. Phone: (615) 322-7026. Fax: (615) 322-7236.

strate a general nuclear localization of this protein in the pancreatic islet, suggesting that this factor is at least present in both α and β cells. We have named the protein encoded by this cDNA clone INSAF (insulin activator factor) and propose that in association with E2A gene products, INSAF can activate insulin gene transcription in pancreatic β cells.

MATERIALS AND METHODS

cDNA library screening. A cDNA library was constructed in λgt11 by Clontech Laboratories, Inc., with oligo(dT)-primed cDNA inserts synthesized from poly(A)-selected RNA prepared from a human pancreatic insulinoma. Bacteriophage infections were carried out according to the methods described by Young and Davis (56). Expression screening was performed using a modification of the procedures described by Vinson et al. (51) and Singh et al. (43), which are detailed below, with a concatenated ^{32}P-labeled double-stranded ICE (−104 5′-TCT GGCCATCTGCTGATCC-3′ −86) probe at a specific activity of approximately 2×10^6 cpm/pmol. Bacteriophage plaques were transferred to isopropyl-β-D-thiogalactopyranoside-impregnated nitrocellulose filters (Schleicher & Schuell). All of the remaining steps in the screening procedure were conducted at 4°C. The filters were immersed in binding buffer (200 mM NaCl, 5 mM $MgCl_2$, 0.5 mM dithiothreitol, 25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.9]) supplemented with 5% (wt/vol) nonfat dry milk. The filters were then gently rocked for 2 h, and the solution was removed and replaced with binding buffer supplemented with 0.25% milk. The filters were transferred after 30 min to 110-mm crystallization dishes (approximately 15 filters per binding reaction mixture) containing 100 ml of the 0.25% milk–binding buffer solution supplemented with 10^6 cpm of ^{32}P-labeled ICE DNA per ml. The filters were left in the binding solution for 2 h and were then washed for an additional 2 h in 0.25% milk–binding buffer. The filters were dried, covered with plastic wrap, and exposed to Kodak X-Omat AR film with an intensifying screen at −70°C. The 6-1 and 4-1 clones were plaque purified after two additional rounds of screening. The 6-1 clone was renamed INSAF.

INSAF DNA sequence analysis. The INSAF cDNA from the 6-1 λgt11 clone was subcloned into the *Eco*RI site of pGEM7 to form the plasmid 6-1 pGEM7. Overlapping exonuclease III deletions of INSAF were generated from 6-1 pGEM7 using the Erase-a-Base system (Promega). The inserts from the deletion clones were size fractionated on agarose gels, and selected clones were sequenced using reverse and/or forward primers. Sequencing was performed using the T7 Sequenase system (U.S. Biochemical Corp.) on double-stranded DNA templates prepared from minipreps.

Preparation of expression plasmids and transient transfections. INSAF expression plasmids were constructed by cloning the INSAF cDNA into the polylinker of the cytomegalovirus (CMV) enhancer-driven expression vector CMV4 (3), in the sense and antisense orientations. The construction of the following HLH expression plasmids used in transfections have been previously described: MyoD (EMC11s [14]), TFE-3 (pSV2A-I3 [4]), Id (E:Id[S] and E:Id[A] [5]), and ITF-1 (pSVE2-5 [23]). ICE-driven expression was assayed from ICE-OVEC, which contains three copies of an oligonucleotide containing rat insulin II gene ICE sequences from −102 to −87 cloned into the *Sal*I site of the OVEC-1 expression vector (55).

The ICE-OVEC expression plasmid was introduced into HIT T-15 2.2.2 (15) and HeLa cells as calcium phosphate coprecipitates made up of the ICE-OVEC plasmid, the expression plasmid or carrier DNA (pUC19), and an internal control plasmid, OVEC-REF (53), as detailed by Whelan et al. (55). Total cellular RNA was isolated by the guanidinium isothiocyanate method, described by Chirgwin et al., 40 to 48 h after transfection (10). Hybridization of the RNA probe to RNA, RNase treatment, and gel electrophoresis were conducted as described previously (55). RNase-protected fragments of 179 and 151 nucleotides in length are expected for correctly initiated ICE-OVEC and OVEC-REF transcripts, respectively. The amounts of correctly initiated transcripts were determined by densitometric scanning of autoradiographs. The amount of transcript detected was normalized to the OVEC-REF internal control signal. Transfections were performed on several occasions with at least two independently isolated plasmid DNA preparations.

In vitro transcription and translation. An INSAF RNA expression plasmid (pET3b/6-1) was made using the pET3b expression plasmid described by Studier et al. (46). pET3b/6-1 was constructed by subcloning the *Eco*NI-*Bam*HI fragment (nucleotide positions 319 to 2622) of 6-1 pGEM7 into *Nde*I-*Bam*HI-cut pET3b. The *Eco*NI and *Nde*I sites were filled in with T4 DNA polymerase. INSAF and E12 RNAs were synthesized in 100-μl reaction mixtures. INSAF RNA was made from 5 μg of pET3b/6-1 using T7 RNA polymerase; E12 RNA was synthesized with T3 RNA polymerase from 5 μg of the E12R plasmid (30). The final RNA product was phenol-chloroform extracted, ethanol precipitated, and dissolved in 20 μl of water. In vitro translation reactions (50-μl mixtures) were performed with 2 μl of RNA, 4 μl of L-[^{35}S]methionine (10 mCi/ml; NEN), and methionine-deficient rabbit reticulocyte lysate (Promega) in accordance with the instructions in the Promega technical manual. Translation reaction mixtures were stored at −70°C.

Preparation of α-INSAF antibodies. The INSAF protein was overexpressed in BL21(DE3)pLysS cells transformed with pET3b/6-1 as described by Studier et al. (46). INSAF was present in the insoluble fraction of the lysed cells. The 80-kDa INSAF protein band was isolated from a preparative 10% sodium dodecyl sulfate (SDS)-polyacrylamide gel. Antibodies to the 80-kDa INSAF protein were raised in rabbits at Bethel Laboratories. Affinity antibodies were purified by INSAF-affinity column purification. A His_6-tagged fusion protein of INSAF was constructed using the pET15b expression vector (Novagen), and the INSAF protein was purified under denaturing conditions using the QIAexpress (Qiagen) protein purification system. An affinity column of purified INSAF was then prepared using an aminolink coupling gel matrix (Pierce). Affinity-pure INSAF antibodies were eluted from this column using a high-salt buffer (5 M LiCl, 10 mM phosphate [pH 7.6]). The INSAF antibodies were dialyzed overnight against phosphate-buffered saline (PBS)-azide (NaN_3 0.05% [wt/vol]) and were concentrated using Centricon-100 miniconcentrators (Amicon). The affinity-pure antibody was then mixed with bovine serum albumin (100 μg/ml), and immunoreactivity and specificity were verified by Western blot (immunoblot) analysis.

DNA-binding assay. A dimer ^{32}P-labeled probe (50,000 cpm/ng; 0.5 ng) containing rat insulin II ICE sequences from −104 to −86 (5′-TCTGGCCATCTGCTGATCCTCTGGCC ATCTGCTGATCC-3′) was incubated with protein extract. Lysed bacterial extracts were prepared from control and pET3b/6-1 (INSAF)-transformed BL21(DE3)pLysS as described elsewhere (46); the INSAF protein was resuspended from the insoluble fraction of cells in 25 mM Tris (pH 8) and 1 mM EDTA. Binding reactions were conducted with 80 μg of bacterial protein containing 20 mM HEPES (pH 7.9), 200 mM

KCl, 4 mM $MgCl_2$, 1 mM EGTA [ethylene glycol-bis(β-aminoethyl ether)-*N,N,N',N'*,-tetraacetic acid], 1 mM dithiothreitol, 4% Ficoll-400, 2 μg of single-stranded calf thymus DNA, and 1 μg of poly(dI-dC). The conditions for the competition analyses were the same, except that the specific and mutant dimer ICE competitor DNAs were included in the mixture (in the amounts detailed in the figure legends) prior to the addition of the protein. Extracts from βTC-3 cells were prepared, and the binding reactions were conducted as described elsewhere (54). Experiments to investigate the effect of the α-INSAF antibodies on the β-cell gel shift complexes were conducted using PBS (pH 7.4)–200 mM KCl–5% glycerol (vol/vol)–100 μg of single-stranded calf thymus DNA per ml–100 μg of poly(dI-dC) per ml. The ^{32}P-labeled probe was added immediately after the addition of the antibody to the extract. The reaction mixtures were incubated for 1 h on ice, and the complexes were resolved by electrophoresis through a 4% polyacrylamide gel.

Immunoblot analysis. The Qiagen nickel column-purified His_6ITF-1 protein was analyzed by Western blot analysis using polyclonal α-E12/E47 antiserum and α-INSAF affinity antibody. The His_6ITF-1 protein was fractionated on a 10% polyacrylamide gel (in the amounts detailed in the figure legends), transferred to Immobilon P, and probed with antibody. INSAF or ITF-1 antibody binding was detected using biotinylated swine anti-rabbit sera and peroxidase-labeled avidin-biotin complex (Dako). The antibody complex was visualized by incubating with the peroxidase substrate 3,3′ diaminobenzidine (DAB; Sigma).

Immunoprecipitation. Immunoprecipitation reactions were performed with 5 μl of reticulocyte lysate-translated ^{35}S-labeled proteins using a modification of the procedures described by Benezra et al. (5). In vitro-translated INSAF and E12R proteins were incubated either together or separately for 1 h at 37°C prior to the addition of antibody. The affinity-purified α-INSAF antibodies and α-E12/E47 antiserum were then added to 150 μl of buffer (10 mM Tris [pH 7.4], 250 mM NaCl, 5 mM EDTA, 0.25% Nonidet P-40) and 20 μl of Affi-Gel protein A (Bio-Rad), and the mixture was incubated first for 30 min at 37°C and then for 1 h at room temperature. The precipitates were washed three times in the buffer described above and were then analyzed on 15% discontinuous SDS-polyacrylamide gels as described previously (27). The gels were fixed in 40% (vol/vol) methanol–7% acetic acid (vol/vol) treated with Entensify (DuPont), dried, and autoradiographed.

Immunocytochemistry. Cells were plated onto poly-L-lysine-coated tissue culture glass slides at 7×10^6 cells per 2-cm^2 slide and fed every 2 days with culture medium. The compositions of the respective culture media are as described for αTC-6 (20), βTC-3 (16), SKNMC (8), and HeLa cell lines (55). RASMC cells were grown in Dulbecco modified Eagle medium containing 10% (vol/vol) fetal calf serum and 50 μg each of penicillin and streptomycin per ml. ATCC CRL 1674 cells were grown in Ham's F12K medium containing 10% (vol/vol) fetal calf serum and 50 μg each of penicillin and streptomycin per ml. Three to five days after plating, the cells were washed twice with L-15 saline (Gibco), fixed with a solution of 4% paraformaldehyde–0.1 M sodium phosphate buffer, and rinsed twice with Tris-saline (0.9% NaCl, 0.1 M Tris [pH 7.4]). The cells were then incubated for 15 min in a solution of 0.3% Triton X-100–0.1% goat serum in Tris-saline and were washed twice with Tris-saline, blocked for 30 min with a 1:30 dilution of the goat serum solution and then incubated for 16 h with either α-INSAF (1:1,000) or α-insulin (1:500 [Linco]) antiserum and then washed twice with Tris-saline. The bound antibody was detected by incubating the dishes for 30 min with a 1:50 dilution of goat-anti-rabbit biotinylated immunoglobulin G, then by incubating with a 1:100 dilution of peroxidase-avidin complex for 30 min, and finally by incubating for 6 min with DAB (0.25 mg of DAB per ml, 0.01% hydrogen peroxide, 0.1 M Tris [pH 7.6]). The sections were dehydrated through graded alcohols after DAB staining and were mounted with Histoclad. The pancreases from adult CD-1 mice (Charles River, Lexington, Mass.) were dissected, and the isolated islets were fixed in 4% paraformaldehyde buffered to pH 7.4 with PBS, embedded, sectioned, and stained with INSAF antiserum (1:1,000) as previously described (35).

RESULTS

Isolation of INSAF. To isolate a cDNA encoding an ICE-binding factor, we screened an oligo(dT)-primed λgt11 cDNA library constructed from the poly(A)-containing RNA isolated from a human pancreatic insulinoma. This library was screened by a modification of existing methods (43, 51) using a multimerized ICE-binding site oligonucleotide as a probe. This isolation approach was undertaken despite the presence of this protein in a heteromeric protein-DNA complex, since individual bHLH proteins can directly bind DNA (29, 30). Two positive clones, designated 6-1 and 4-1, were isolated from a primary screen of approximately 750,000 phage plaques. The binding specificities of 6-1 and 4-1 were determined by screening plaque lifts using multimerized wild-type and mutant ICE probes. Restriction endonuclease cleavage patterns and partial nucleotide sequence analysis of the 4-1 cDNA insert indicated that its sequence was identical to that of 6-1 (data not shown). The 6-1 recombinant was chosen for further analysis. The 6-1 clone was renamed INSAF. All of the results described below were conducted with this INSAF cDNA.

The INSAF clone contained a 2,622-bp insert. The largest open reading frame within this clone could encode a protein of 744 amino acids (Fig. 1). However, there were no in-frame termination codons and no consensus polyadenylation signal within this reading frame. This clone is therefore presumed to be missing a portion of its 3′-coding sequences. There were 390 bp of 5′-noncoding sequence within the INSAF clone. Overall, the INSAF polypeptide is hydrophilic in nature; however, there is a high concentration of hydrophobic amino acids within its amino-terminal region. Almost two-thirds of the amino acid residues within the first 110 amino acids of the protein are nonpolar. INSAF contains no cysteines. There are several potential sites of phosphorylation by cyclic AMP-dependent kinase (amino acids 109 and 187), calmodulin kinase II (amino acids 109 and 187), protein kinase C (amino acid 334), glycogen synthase kinase-3 (amino acids 71, 109, 191, 201, 234, 246, 312, 582, and 634), and tyrosine kinase (amino acid 566) within the INSAF polypeptide. Using the method described by Altschul et al. (2), we compared our protein and amino acid sequences with those in the various databases at the NIH Genebank service and showed that INSAF was a unique protein with two potential structural motifs: a leucine zipper (between amino acids 1 and 61) and a bHLH (between amino acids 214 and 280).

INSAF protein is capable of binding specifically to ICE sequences. To express the INSAF protein in *Escherichia coli*, the coding region from the 2.6-kb INSAF cDNA was subcloned in frame into the pET3b vector described by Studier et al. (46). Translation of the full-length INSAF cDNA in bacteria resulted in one major protein product of approximately 80 kDa. The bacterially synthesized INSAF protein was examined for binding to an ICE-radiolabeled probe by the electrophoretic mobility shift assay. A single major protein-DNA

```
                                                CGGTACAAATTCGACAACGGTAAAACAGCATTTGA   35
AACAGAATTAGAACGTATGCTTTACTCAATCAGTGACGCAGGCTTTGAAATTCCAGACGTTGATTTTACTT   106
TCGACAACATTGACCAATTCTTGAAGTCAAAGCAATATAAGGCATTCATTCGAGTTCGTCCACAAAAGAAC   177
AATCCAGAATACACTGAAGCTATTTTCATCTCAAAAGAACAATTCGACAAAGAAATCGAAAAACAAGCAAA   248
TGACGCATTTGGTGACGTAGTTGACCCATTCGCAACAGCTACACCAGTAACAGACAACCCATTTGCTTTAAT   319
AAATGCCCTTTTATTATAGGGCGTAAATAATCGTGCGCTATCCACGTTAAAAGGGTAAGAAAAGGAGTATTATC   390

 M  N  T  K  K  L  A  V  T  N  G  V  V  G  L  V  G  G  I  A  L  L  L  G   24
ATGAACACAAAGAAGTTAGCAGTAACAAATGGAGTAGTAGGACTTGTCGGTGGAATCGCACTGTTGTTAGGT   462

 G  W  I  V  L  Y  S  G  N  R  W  N  R  T  F  A  I  V  N  V  L  F  L  L   48
GGTTGGATCGTATTGTATTCAGGCAACAGGTGGAACAGGACGTTCGCAATTGTAAATGTGTTATTCTTATTA   534

 L  K  I  A  T  L  V  L  G  I  V  G  L  V  Q  F  S  K  E  A  V  I  T  K   72
TTGAAAATCGCAACGCTTGTTCTTGGTATTGTAGGACTTGTACAATTTAGCAAGGAAGCAGTTATCACAAAG   606

 A  P  S  V  L  L  V  V  G  G  A  L  A  I  I  P  F  M  G  W  I  G  G  I   96
GCACCTAGCGTATTACTAGTTGTCGGCGGTGCATTGGCAATTATCCCATTCATGGGGTGGATCGGTGGTATT   678

 L  S  I  I  G  G  S  I  Y  L  A  I  S  R  N  L  S  N  K  M  A  F  I  A  120
CTTTCAATCATCGGTGGTTCAATCTACCTAGCAATATCAAGAAATTTAAGTAACAAAATGGCGTTTATCGCC   750

 V  H  I  K  K  G  D  H  D  M  I  N  N  S  Y  L  L  K  A  D  L  K  R  E  144
GTACATATTAAAAAAGGAGACCATGACATGATTAACAATTCGTATCTATTAAAGGCAGATTTAAAGCGTGAA   822

 K  D  I  P  L  I  S  R  E  S  I  I  K  L  A  K  A  H  E  R  G  D  V  R  168
AAGGACATTCCTTTAATCAGCCGTGAAAGTATTATCAAATTAGCTAAGGCACACGAACGTGGCGATGTTAGA   894

 Y  L  N  E  A  E  T  N  V  Y  D  L  L  T  K  I  E  V  K  R  L  A  S  L  192
TATTTAAATGAAGCAGAAACTAATGTATACGACTTACTGACAAAAATAGAAGTAAAGCGGTTAGCTAGTCTG   966

 Y  K  S  F  I  F  D  F  S  I  E  D  S  T  V  S  Y  T  F  A  D  I  E  E  216
TACAAGTCATTTATTTTTGATTTCTCAATTGAAGATAGTACAGTGTCATACACTTTCGCAGATATTGAAGAA   1038

 A  I  R  D  N  I  G  V  K  N  K  R  L  L  E  D  S  T  N  L  M  S  D  I  240
GCTATACGTGATAATATCGGTGTCAAAAACAAGCGTTTGCTAGAAGATAGTACTAACCTTATGAGTGATATT   1110

 I  R  F  A  V  T  P  V  D  Y  I  F  D  F  H  K  N  D  N  G  N  L  E  V  264
ATCAGATTCGCAGTGACACCAGTTGATTATATATTTGATTTTCATAAAAATGACAATGGCAATTTAGAAGTA   1182

 Q  R  D  Q  V  L  E  R  F  I  E  V  L  E  Q  Q  T  Y  E  N  V  W  S  D  288
CAACGTGATCAAGTATTAGAACGATTTATCGAAGTTCTTGAACAGCAGACGTATGAAAACGTATGGTCTGAC   1254

 L  M  K  Y  P  F  D  N  Q  T  H  G  I  R  K  H  L  M  L  G  V  A  A  S  312
TTAATGAAATATCCATTTGACAACCAGACACACGGGATACGCAAACATTTAATGCTCGGTGTCGCAGCGTCA   1326

 T  G  K  S  I  I  L  E  A  I  G  K  L  Y  E  A  S  A  S  F  K  S  N  R  336
ACAGGTAAATCTATTATCCTTGAAGCTATTGGTAAGTTATACGAAGCGTCAGCTTCATTCAAATCAAACAGA   1398

 P  E  N  G  F  D  A  A  N  W  N  A  D  I  I  D  K  F  V  V  L  L  D  D  360
CCTGAAAACGGCTTTGACGCAGCAAATTGGAACGCTGACATTATTGATAAGTTTGTTGTATTGCTTGATGAT   1470

 D  N  P  N  V  S  I  N  E  D  Y  L  K  N  F  L  S  N  N  M  P  Q  Q  L  384
GATAACCCAAACGTCTCAATCAACGAAGACTACCTGAAAAACTTCCTAAGCAATAATATGCCACAACAATTA   1542

 A  R  G  G  G  Q  R  W  S  E  I  Y  Y  G  S  S  V  I  A  L  N  T  A  P  408
GCACGCGGTGGCGGTCAACGTTGGTCAGAAATTTATTACGGGTCATCGGTTATCGCTTTAAACACAGCACCA   1614

 D  F  L  K  N  Q  Q  N  N  K  R  V  I  F  L  K  L  T  Q  S  I  E  E  M  432
GACTTCCTTAAAAACCAACAAAACAATAAGCGTGTCATTTTCTTGAAGTTGACCCAGTCAATTGAAGAAATG   1686

 F  T  S  E  E  L  N  H  L  H  N  L  E  P  K  H  I  L  G  Y  V  N  S  E  456
TTCACGTCAGAAGAATTAAATCATTTACACAATTTAGAGCCTAAGCACATCTTGGGTTACGTCAATTCAGAA   1758

 P  T  K  L  F  K  W  R  N  D  W  S  G  V  I  E  D  L  K  E  D  V  K  Q  480
CCAACAAAGTTGTTTAAGTGGCGTAATGACTGGTCAGGCGTTATTGAAGACTTGAAAGAAGATGTTAAGCAA   1830

 H  I  S  M  R  G  Y  A  S  L  N  D  L  K  K  T  F  G  D  K  V  M  R  D  504
CACATCTCAATGCGTGGTTACGCAAGCCTTAATGACCTTAAAAAGACGTTTGGCGACAAAGTTATGCGTGAC   1902

 Y  Q  R  T  V  E  K  A  K  N  I  T  V  N  G  E  A  F  Y  G  L  R  Q  D  528
TATCAACGAACGGTTGAAAAGGCTAAAAACATCACAGTAAACGGTGAAGCGTTCTACGGTTTAAGACAAGAC   1974

 A  I  I  D  E  P  Y  R  V  D  Q  L  T  F  S  A  F  I  N  L  H  D  V  I  552
GCAATCATTGATGAGCCTTATCGTGTCGATCAACTAACGTTTAGTGCATTTATCAACTTACATGACGTGATT   2046

 G  E  R  N  F  E  S  L  D  E  F  K  N  D  I  E  E  Y  G  K  L  K  K  E  576
GGCGAACGAAATTTGAAAGCCTTGACGAGTTTAAAAACGATATTGAAGAATACGGAAAGTTGAAGAAAGAA   2118

 E  Q  P  L  F  S  P  A  V  Y  H  D  E  L  T  K  A  N  L  T  G  F  V  G  600
GAGCAACCACTATTCAGTCCAGCTGTTTATCATGACGAGTTGACCAAAGCAAATCTAACAGGCTTTGTTGGC   2190

 V  V  L  D  I  D  G  A  N  A  A  D  M  D  E  L  E  D  K  I  A  M  T  G  624
GTTGTTCTGGATATTGACGGTGCAAACGCAGCAGACATGGACGAGCTAGAAGATAAAATCGCTATGACTGGT   2262

 I  S  A  V  A  W  E  T  A  S  S  T  P  S  S  L  R  A  R  I  W  F  Y  R  648
ATATCAGCAGTTGCTTGGGAAACGGCATCAAGCACACCGTCATCGTTGAGAGCGCGTATTTGGTTCTATCGT   2334

 V  G  G  G  L  E  E  Y  A  S  L  V  K  K  I  A  M  K  I  D  E  P  F  D  672
GTCGGTGGTGGTTTGGAAGAATATGCTAGTCTAGTTAAAAAGATTGCGATGAAGATTGACGAGCCATTTGAC   2406

 R  A  S  M  P  A  E  H  R  F  F  I  G  G  T  N  V  R  L  I  N  T  D  S  696
CGTGCTAGCATGCCAGCTGAACATCGTTTCTTCATTGGTGGTACAAATGTACGCCTAATCAATACAGACAGC   2478

 T  A  E  N  K  P  M  T  D  M  K  K  R  G  L  V  R  A  V  Q  N  A  K  720
ACAGCGCTGAAAATAAGCCAATGACAGACATGAAGAAGCGTGGGCTAGTCCGTGCAGTTCAGAATGCTAAG   2550

 D  G  E  R  E  Q  I  T  Y  W  A  L  M  R  A  E  Q  E  N  G  G  K  E  F  744
GACGGCGAACGTGAGCAGATAACTTACTGGGCATTAATGCGAGCAGAGCAAGAAAACGGTGGAAAAGAATTC   2622
```

FIG. 1. Nucleotide sequence of INSAF and its predicted amino acid sequence. The nucleic acid residues are numbered from the nucleotide in the cDNA, and the amino acid residues are numbered from the beginning of the long open reading frame.

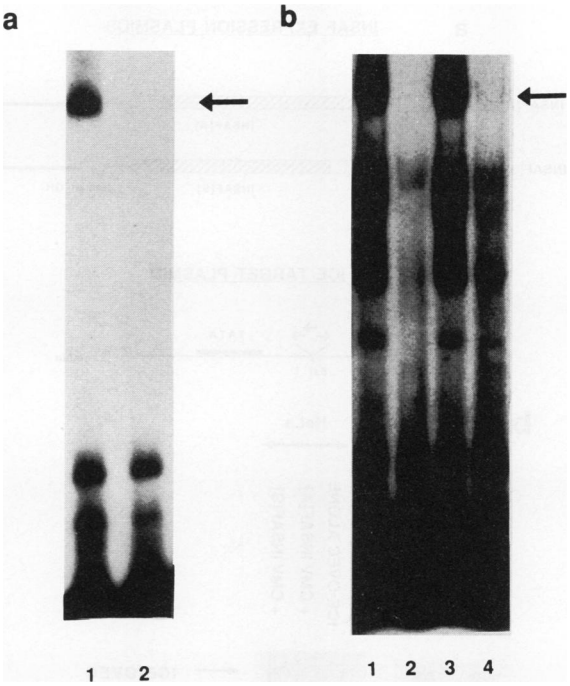

FIG. 2. Binding of bacterially synthesized INSAF to ICE sequences. (a) Binding of the INSAF protein was assayed from 80 μg of *E. coli* extracts by the gel shift assay with a ICE probe containing two copies of the rat insulin II gene sequences from −104 to −86. Lanes: 1, extract prepared from INSAF-expressing cells; 2, control extract. The appearance of the INSAF-specific complex was also detected in less than 10 μg of extract. (b) Competition analysis. Binding of the bacterially synthesized INSAF protein to the wild-type ICE dimer probe was analyzed at a 50-fold molar excess of unlabeled competitor to probe. The competitors contained a single-site noncomplementary transversion mutation at either bp −100 or bp −93. Competitors had the same nucleotide length as the wild-type probe. Lanes: 1, no competitor; 2, wild-type competitor; 3, mutant −93 competitor; 4, mutant −100 competitor. The specific INSAF-DNA complex is denoted by the arrow, and the autoradiograph in panel b is overexposed to show the absence of this specific complex. The faster-migrating complexes are nonspecific bacterial complexes.

band was uniquely found in INSAF-expressing cell extracts (Fig. 2a). The binding specificity of the protein(s) within this complex was examined by analyzing the effect of wild-type and mutant ICE competitor DNAs on complex formation. Previously, we demonstrated that point mutations within the ICE bHLH binding site at bp −98, −97, −94, and −93 prevented activator binding in vitro and function in vivo, whereas mutations outside this region had little or no effect on activity (54). Similarly, DNA containing the bp −93 mutation did not compete with the wild-type probe for the unique protein-DNA complex in INSAF-expressing cell extracts, whereas DNA containing the bp −100 mutation, a functionally silent mutant, competed as effectively as the wild-type competitor (Fig. 2b). DNA containing the bp −97 mutation also did not compete (data not shown). The properties of INSAF protein binding in these competition experiments indicate that it interacts with the ICE with a specificity similar to that of the endogenous activator.

INSAF activates ICE-mediated expression. To determine whether INSAF encoded a transcription factor that could activate ICE-mediated expression, we analyzed the effect of cotransfection of INSAF sequences upon the expression of

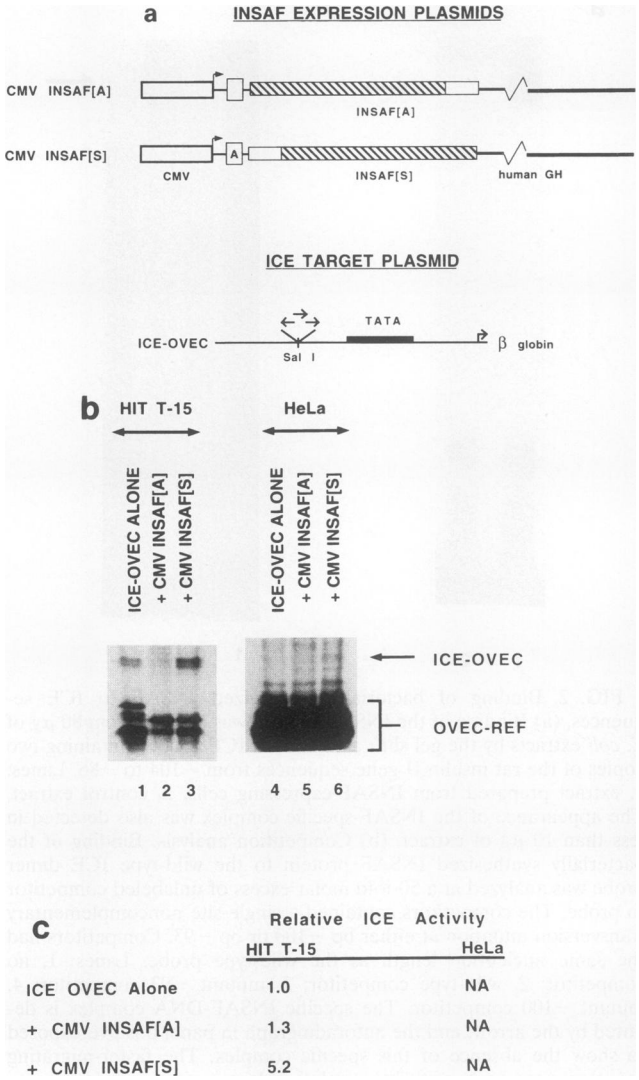

FIG. 3. β-cell-specific activation of ICE-driven activity by INSAF. (a) Structures of INSAF and ICE expression plasmids. Human INSAF cDNA sequences were cloned in both the sense (CMV INSAF[S]) and antisense (CMV INSAF[A]) orientations into the plasmid pCMV4 (3) as described in Materials and Methods. The open and hatched boxes denote untranslated and translated INSAF mRNA sequences, respectively. CMV, the immediate-early promoter region of the human CMV; A, a segment of the alfalfa mosaic virus 4 RNA that contains a translational enhancer; human GH, human growth hormone gene sequences that encode transcriptional termination and polyadenylation signals. ICE-OVEC contains three copies of an oligonucleotide containing rat insulin II gene ICE sequences from −102 to −87 (see references 11 and 55). The orientations of the ICE inserts are indicated by the arrows. (b) Autoradiogram of RNase protection analyses. HIT T-15 and HeLa cells were transfected with 2 μg of ICE-OVEC, either 8 μg of CMV INSAF[S] or 8 μg of CMV INSAF[A], and 0.5 μg of an internal control plasmid, OVEC-REF. RNA levels were analyzed as described in Materials and Methods. The signals for the correctly initiated ICE-OVEC and OVEC-REF transcripts are shown. (c) Fold activation of ICE-driven transcription by INSAF. Relative amounts of correctly initiated transcripts were determined by densitometric scanning of the autoradiogram and were normalized relative to the OVEC-REF internal control signal. Fold activation is calculated as the ratio of densitometric units of signal produced from ICE-OVEC in the presence of INSAF divided by the units of signal produced in the absence of INSAF. NA, not applicable.

ICE-OVEC, a reporter plasmid which contains three copies of the ICE inserted directly upstream of the rabbit β-globin TATA box and coding sequences. We previously showed that expression of ICE-OVEC is pancreatic β cell specific and entirely dependent on the ICE (55). The INSAF cDNA was cloned into the CMV enhancer-driven expression plasmid CMV4, in both the sense and antisense orientations (Fig. 3a). The INSAF mRNA expressed from this plasmid contains at its 5′ end a segment of the alfalfa mosaic virus 4 RNA that encodes a translational enhancer and at the 3′ end the transcription termination and polyadenylation signals from the human growth hormone gene.

ICE-OVEC was introduced into insulin-producing (e.g., HIT T-15 [15]) and non-insulin-producing (e.g., HeLa) cell lines by calcium phosphate coprecipitation with or without the INSAF expression plasmids. ICE expression was normalized to expression from a cotransfected simian virus 40 enhancer-driven rabbit β-globin expression plasmid, OVEC-REF. The results of a typical experiment are shown in Fig. 3b. Low levels of activity are obtained from HIT T-15 cells transfected with ICE-OVEC in the absence of CMV INSAF. This level of ICE-OVEC expression is probably due to the endogenous *trans* activator present in this β cell line (55). A four- to fivefold increase in ICE-driven β-globin mRNA levels was observed in INSAF[S] transfected HIT T-15 cells relative to those that were transfected with ICE-OVEC alone or with INSAF[A] (Fig. 3c). In contrast, INSAF[S] had no effect on ICE-OVEC activity in HeLa cells. Similar results were obtained upon transfection of these sets of plasmids into one other insulin-producing (βTC-1 cells [16]) and two non-insulin-producing cell types (baby hamster kidney and NIH 3T3 cells) (data not shown). To further address the question of the specificity of the INSAF effects, CMV INSAF[S] was cotransfected with mutant ICE-reporter plasmids. ICE mutations that inhibited the ability of the cellular activator to stimulate ICE transcription in β cells also prevented INSAF-mediated activation, whereas INSAF stimulated transcription driven from silent ICE mutation expression plasmids (data not shown). We conclude from these results that INSAF can specifically interact and activate ICE-mediated transcription in β cells.

We considered that the inability of INSAF to *trans* activate ICE-mediated expression in non-insulin-producing cells may be due to a negative regulator(s) of ICE activity (55). To test whether the repression was specific to INSAF activation or of a more general type, we analyzed the effect of a number of bHLH activators on ICE-mediated expression. ICE-OVEC was transfected into HeLa cells, either alone or with plasmids expressing the following bHLH proteins: MyoD (a skeletal muscle-specific activator [14]), ITF-1 (a ubiquitously distributed E47-like factor [23]), and TFE3 (a generally distributed factor that contains leucine zipper motifs adjacent to its bHLH region [4]). All of these factors bind to ICE sequences in vitro (12, 54). As shown in Fig. 4, ICE-mediated expression was stimulated in HeLa cells cotransfected with the MyoD and ITF-1 expression plasmids. MyoD and ITF-1 can also activate ICE-OVEC expression in HIT T-15 cells (data not shown). Although MyoD is uniquely expressed in skeletal muscle cells, this protein has been shown to activate expression in nonmyogenic cell types in transient expression assays (50). No activation was observed from either the INSAF or the TFE3 protein. The inability of TFE3 to activate may reflect its relatively weak binding to the ICE in vitro (data not shown). It is not yet known why the stimulatory activity mediated by INSAF in β cells is blocked in HeLa cells. However, it appears to be a general property of non-insulin-producing cells, since ICE-mediated expression was not activated by INSAF when assayed

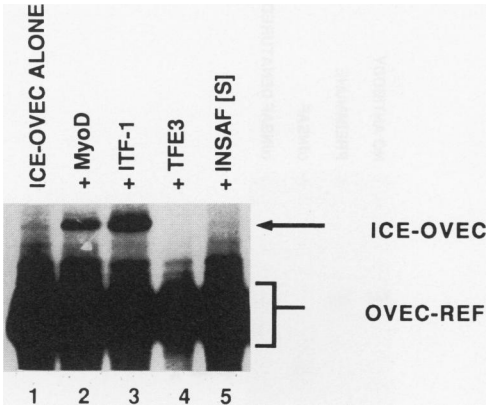

FIG. 4. Effects of coexpression of MyoD, ITF-1, and TFE3 in HeLa cells on ICE-driven transcription. HeLa cells were transfected with 2 μg of ICE-OVEC, 0.5 μg of OVEC-REF, and 8 μg of either of the following expression plasmids: MyoD (EMC11s [14]), ITF-1 (pSVE2-5 [23]), or TFE3 (pSV2A-λ3 [4]). OVEC-derived RNA levels were analyzed by RNase protection analyses. The signals for the correctly initiated ICE-OVEC and OVEC-REF transcripts are shown.

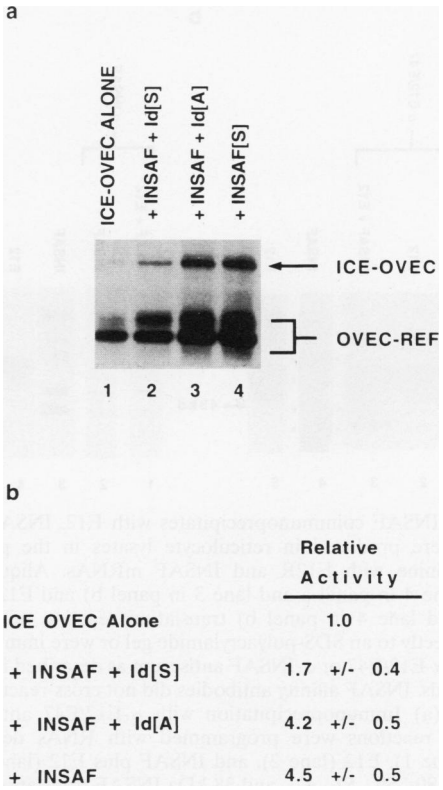

FIG. 5. Id inhibits INSAF-dependent *trans* activation. HIT T-15 cells were cotransfected with 2 μg of ICE-OVEC, 5 μg of CMV INSAF[S], 0.5 μg of OVEC-REF, and 5 μg of an expression plasmid with the Id cDNA cloned in either the sense (Id[S]) or antisense (Id[A]) orientation. The signals for the ICE-OVEC and OVEC-REF transcripts were determined by RNase protection analyses. Relative amounts of correctly initiated transcripts were determined by densitometric scanning of the autoradiogram and were normalized relative to the OVEC-REF internal control signal. The data are means ± standard errors relative to the sample with ICE-OVEC alone from four separate experiments.

in other mammalian cell lines that do not produce insulin (data not shown).

Id inhibits *trans* activation by INSAF. The stimulation of ICE-mediated expression by the endogenous β-cell activator can be inhibited by Id, a negative regulator of bHLH function (11). This HLH protein lacks a basic region but can form stable protein-protein heterodimers with E12 and E47/ITF-1 through its HLH domain (5). However, since Id lacks a basic DNA-binding region, these heterodimers are unable to bind DNA and activate transcription. To determine the effect of Id on INSAF-dependent activation, HIT T-15 cells were cotransfected with the ICE-OVEC reporter plasmid, CMV INSAF[S], and either an Id sense (E:Id[S]) or an Id antisense (E:Id[A]) expression plasmid (Fig. 5). INSAF alone stimulated ICE-OVEC expression approximately fourfold. This activity was unaffected by cotransfecting the E:Id[A] expression plasmid along with CMV INSAF[S]. However, cotransfections with the E:Id[S] plasmid significantly repressed INSAF-mediated activity. These results indicate that Id inhibits the activity of INSAF. We believe that this is a consequence of direct competition between Id and INSAF for limiting amounts of E2A gene products present in these cells. These results are consistent with the possibility that the ICE activator is a heteromer of INSAF and these E proteins.

INSAF and E12 form a heteromeric complex. To determine whether INSAF and E12 can form heteromers, immunoprecipitation assays were performed with antibodies to the INSAF or the E2A proteins. These antibodies were used to immunoprecipitate ^{35}S-labeled proteins from reticulocyte lysates programmed with E12 and INSAF mRNAs. RNAs prepared from E12 and INSAF cDNAs were separately translated in lysate containing [^{35}S]methionine. The products were analyzed either directly on SDS-polyacrylamide gels or after immunoprecipitation with α-INSAF or α-E12/E47 antiserum.

Several labeled protein products were obtained upon translation of INSAF RNA, possibly because of premature termination during in vitro translation (Fig. 6). The major polypeptides are 80, 59, 52, 42, and 38 kDa. E12 RNA directed the synthesis of a major polypeptide species of 50 kDa, which is the bona fide E12 translation product; the 68-kDa protein is presumably a fusion protein of E12 and plasmid-coding sequences. E12 and INSAF proteins are readily resolvable on SDS-polyacrylamide gels. The antibody recognizing the E2A proteins precipitated the 80-, 59-, 52-, 42-, and 38-kDa INSAF translation products in an E12-dependent fashion when the two translation reactions were combined (Fig. 6a; compare lane 3 with lanes 2 and 4). This antiserum is specific, as demonstrated by the inability of α-E12/E47 antiserum to immunoprecipitate the INSAF RNA translation products directly. In addition, the α-E12/E47 preimmune serum did not recognize either E2A gene products or INSAF (data not shown). Heteromerization between the INSAF translation products and the 50-kDa E12 protein was also detected using affinity-purified antiserum to INSAF (Fig. 6b). The amounts of E12 and INSAF that were immunoprecipitated in these reactions were low, suggesting that protein-protein dimerization in vitro between E12 and INSAF is relatively inefficient. However, these results also indicate that INSAF can heteromerize with E12 and are entirely consistent with the reports demonstrating that proteins antigenically related to the E2A gene products are contained within the ICE activator complex (11, 18, 42).

INSAF is a component of the ICE-activator factor complex. To determine whether INSAF-like proteins are part of the ICE

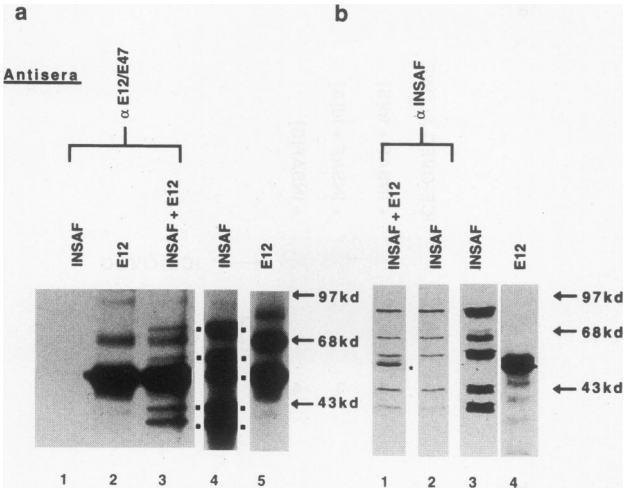

FIG. 6. INSAF coimmunoprecipitates with E12. INSAF and E12 proteins were produced in reticulocyte lysates in the presence of [^{35}S]methionine with E12R and INSAF mRNAs. Aliquots of the INSAF (lane 4 in panel a and lane 3 in panel b) and E12 (lane 5 in panel a and lane 4 in panel b) translation reaction mixtures were applied directly to an SDS-polyacrylamide gel or were immunoprecipitated with α-E12/E47 or α-INSAF antiserum as described in Materials and Methods. INSAF affinity antibodies did not cross-react with ITF-1 (Fig. 8b). (a) Immunoprecipitation with α-E12/E47 antibody. The translation reactions were programmed with RNAs derived from INSAF (lane 1), E12 (lane 2), and INSAF plus E12 (lane 3). Boxes denote the 80-, 59-, 52-, 42-, and 38-kDa INSAF translation products. (b) Immunoprecipitation with α-INSAF affinity antibody. The translation reactions were programmed with INSAF plus E12 (lane 1) and INSAF (lane 2) RNAs. Asterisk, the 50-kDa E12 translation product.

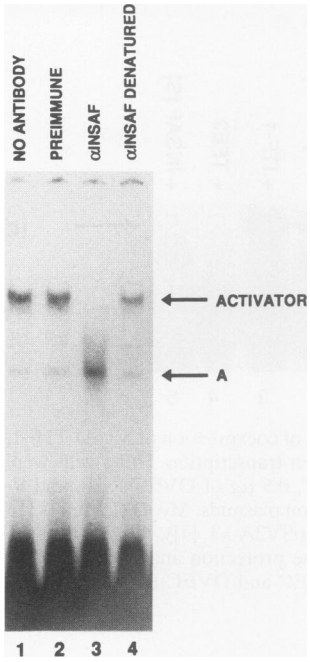

FIG. 7. Affinity-purified α-INSAF antibodies can reduce the formation of the ICE activator gel shift complex. DNA binding analysis was performed with βTC-3 cell extracts as detailed in Materials and Methods. Lanes: 1, extract alone; 2, extract incubated with preimmune antibody; 3, extract incubated with α-INSAF antibody; 4, extract incubated with denatured α-INSAF antibody.

activator complex in β cells, we tested the effect of INSAF-specific antiserum on the gel mobility shift complexes formed with β extracts. Binding experiments were performed with a ^{32}P-labeled wild-type ICE probe and extracts prepared from βTC-3 cells, a mouse pancreatic β cell line (16). There are two ICE-binding complexes formed under these conditions, corresponding to the activator and the A complex (54). The activator complex appears to be restricted to pancreatic α and β cell lines (38), whereas the A complex is more generally distributed (54). Incubation with anti-INSAF affinity antibody reduced the levels of the ICE activator complex (Fig. 7; compare lanes 1 and 3). By contrast, this antiserum had no effect on the formation of the generally distributed A complex (Fig. 7) (54). However, neither preimmune nor denatured INSAF antibody preparations affected the levels of activator complex (Fig. 7). The reaction of the INSAF affinity antibodies with the ICE activator complex could also be demonstrated by running a β-cell extract preparation over an immobilized α-INSAF antibody affinity column. Under these conditions, the INSAF antibody specifically retained the ICE activator complex present in the βTC-3 extracts loaded onto the column (data not shown).

The ability of our INSAF affinity antibody to abolish formation of the ICE activator may have resulted from cross-reaction between the INSAF antibodies and the ubiquitously expressed E2A proteins (e.g., E12, E47, or ITF-1) present in this complex. Each member of the E2A family of proteins shares regions of homology, especially within their carboxy terminal bHLH domains (23, 30). To determine whether INSAF antibodies recognized epitopes in these proteins, a full-length His_6-tagged ITF-1 fusion protein was synthesized in bacteria

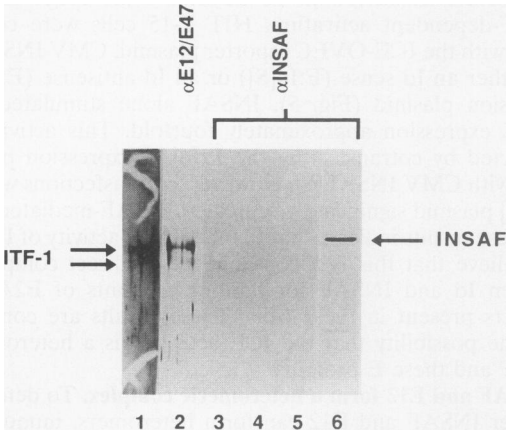

FIG. 8. α-INSAF antibodies do not react with the ITF-1 protein. Western blot showing that the α-INSAF antibody does not cross-react with the ITF-1 bHLH protein. Affinity-purified His_6ITF-1 protein was fractionated on a 10% polyacrylamide gel, transferred to Immobilon P, and probed with antibody. Lanes: 1, silver staining of 3 μg of affinity-purified His_6ITF-1 protein; 2, 3 μg of affinity-purified His_6ITF-1 protein reacted with antiserum to E12/E47; 3, 4, and 5, 3, 7.6, and 15 μg, respectively, of affinity-purified His_6ITF-1 protein reacted with the affinity-purified INSAF antiserum; 6, crude bacterial lysate containing the INSAF protein (22 μg) reacted with antiserum to INSAF. The major product corresponds to the correctly sized 66-kDa His_6ITF-1 protein; however, the smaller products are also detected with the antiserum to E12/E47. The α-INSAF antibody reacts in crude bacterial lysates with the 80-kDa INSAF protein (labeled). The lower-molecular-mass protein is due to a cross-reaction between the second antibody and an unknown protein in the bacterial extract.

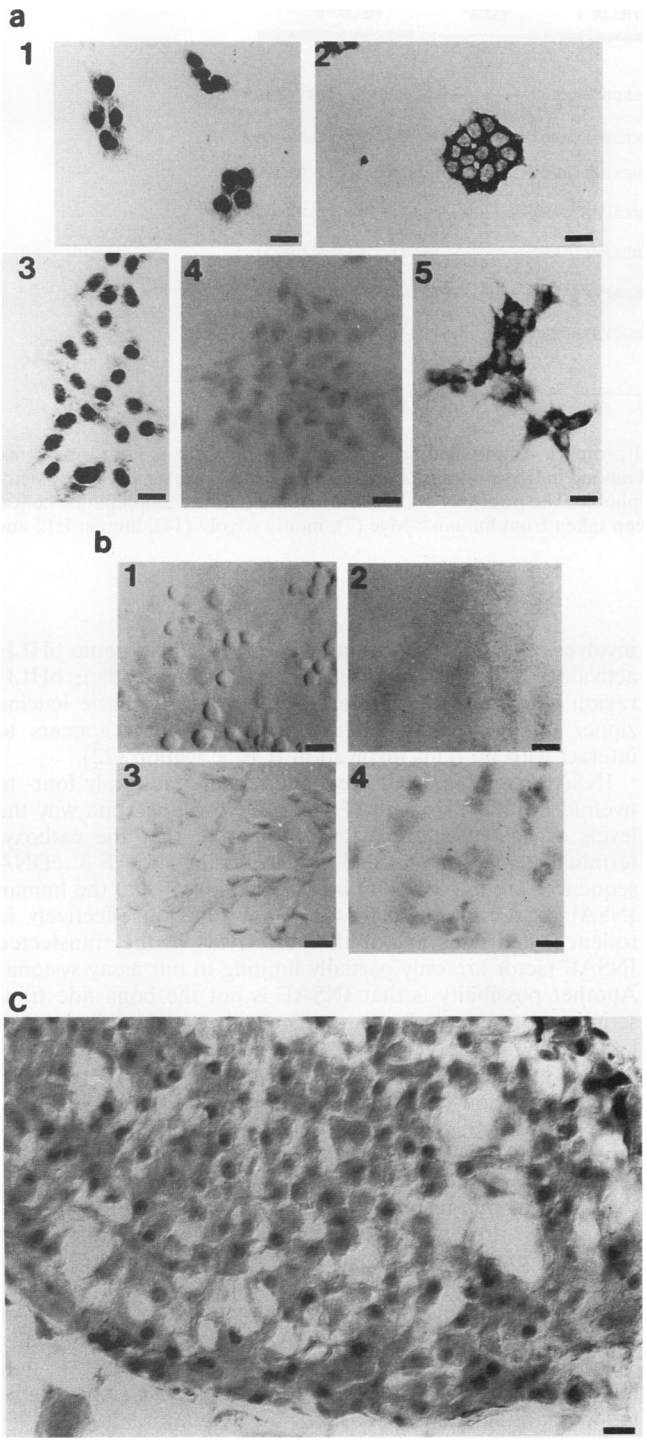

FIG. 9. Immunostaining analysis of INSAF expression. (a) INSAF is localized in the nuclei of pancreatic α and β cells. 1, βTC-3 cells immunostained with α-INSAF antiserum. Note specific nuclear staining. 2, βTC-3 cells immunostained with insulin antiserum. The photomicrograph illustrates the localization of immunoreactive product in the cytoplasm. 3, 4, and 5, αTC-6 cells immunostained with α-INSAF affinity, insulin, and glucagon antisera, respectively, demonstrating nuclear staining to INSAF protein in a non-insulin-expressing cell type. Bar, 10 μm. (b) Various nonislet cell types lack INSAF immunostaining. HeLa cells (panel 1), RASMC aorta smooth muscle cells (panel 2), SKNMC neuroblastoma cells (panel 3), and CRL-1674 pancreatic exocrine cells (panel 4) were incubated with α-INSAF antiserum. Bars, 8 μm for panel 2 and 10μm for panels 1, 3, and 4. (c) Photomicrograph illustrates the immunohistolocalization of INSAF in islet cells of adult mouse pancreas. Bar, 20 μm. Note that the immunoreactive staining is localized to the nuclei of islet cells and that the exocrine cells surrounding the islet are unstained.

and purified by nickel column affinity chromatography. The only protein-staining bands detected after chromotography reacted with αE12/E47 antiserum, which indicated that this protein was essentially purified to homogeneity (Fig. 8). However, there was no detectable recognition of the ITF-1 protein when up to 15 μg of the purified protein was assayed with α-INSAF antiserum by Western blot analysis (Fig. 8). In contrast, the α-INSAF antiserum reacted to the 80-kDa INSAF protein (Fig. 8). INSAF antisera also did not recognize epitopes on the E2A (E47 or ITF-1) proteins present in E2A-ICE-binding complexes (data not shown). Thus, we conclude that the α-INSAF affinity antibodies recognize the ICE activator complex as a result of the recognition of epitopes unique to the INSAF protein. These results appear to indicate that INSAF or an antigenically related protein is found within the ICE activator gel shift complex.

INSAF is a nuclear protein expressed only in pancreatic islet α and β cells. The expression and localization of the INSAF protein were examined by immunohistochemical analyses using various cell lines, including a mouse pancreatic islet β cell line, βTC-3 (16); a mouse pancreatic islet α cell line, αTC-6 (20); a rat pancreatic exocrine cell line, ATCC-CRL1674; a rat aorta smooth muscle cell line, RASMC (41); a human neuroblastoma cell line, SKNMC (8); and HeLa cells. The INSAF protein was detected only in the pancreatic islet α and β cell lines (Fig. 9a and b). Furthermore, the staining with antiserum to INSAF was localized within the nuclei of βTC-3 and αTC6 cells, whereas insulin and glucagon staining was detected only in the cytoplasm of these cells (Fig. 9a and b). INSAF-specific nuclear staining was also detected in HIT T-15 cells (data not shown). The levels of the ICE activator present in a mobility shift experiment using HIT and αTC-6 cell extracts were also affected by the α-INSAF antiserum (data not shown). INSAF nuclear staining was also detected in isolated mouse pancreatic islets (Fig. 9c). INSAF immunostaining was uniformly distributed within the islet but was absent from the acinar cells surrounding the islet. Since the islet is predominately composed of α (approximately 20%) and β (approximately 75%) cells, these results indicate that INSAF is a nuclear protein present in at least two developmentally related islet cell types, α and β cells.

DISCUSSION

Pancreatic β-cell-type-specific expression of the insulin gene depends on regulation mediated by its enhancer. There are a number of distinct DNA sequence elements within the enhancer that are important for expression in β cells (for a review, see reference 44). Characterization of the insulin gene enhancer region indicates that cell-type-specific transcription can be recapitulated in cell culture with reporter constructs driven by subregions of the insulin enhancer. These regions, which lie between nucleotides −247 and −197 in the rat insulin I gene and nucleotides −126 and −86 bp in the rat insulin II gene, contain sequence motifs in addition to the ICE that act in synergy to give a β-cell-specific enhancer effect. These same *cis*-acting sequences also appear to mediate, at least in part, the induction of insulin gene transcription by glucose (19, 40). These results suggest that selective and inducible transcription

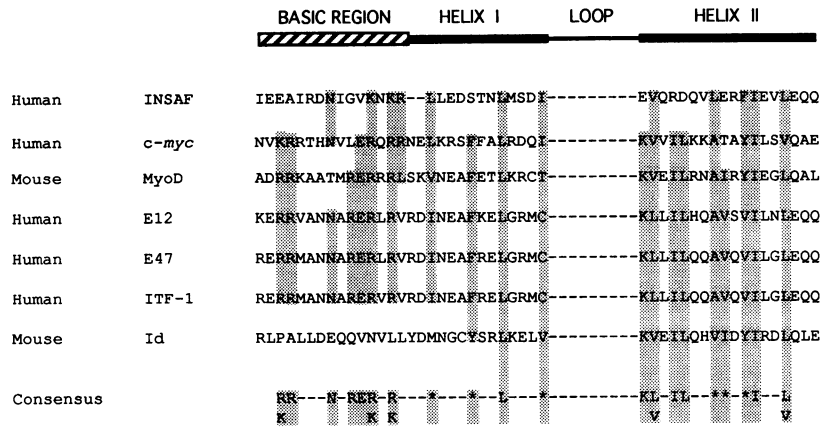

FIG. 10. Amino acid sequence comparison of INSAF with other HLH proteins. Amino acid sequence alignments were made to maximize homology within the bHLH region. The shaded amino acids are conserved among bHLH proteins. Amino acids D and E are acidic, amino acids R and K are basic, and amino acids A, C, V, I, L, M, F, Y, and W are hydrophobic. The positions of the basic and the putative amphipathic helical regions within these proteins are indicated (14, 29, 33). Sequences have been taken from human c-Myc (7), mouse MyoD (14), human E12 and E47 (29), human ITF-1 (23), and mouse Id (5).

of the insulin gene is regulated by a protein that binds to the ICE sequence. In the present study, we have cloned and characterized the gene (cDNA) encoding INSAF, which appears to act as a positive regulator of ICE-mediated activity.

The INSAF protein was isolated from a human insulinoma cDNA library as a consequence of its ability to bind specifically to a concatamerized ICE probe. We have shown that INSAF stimulated ICE-driven expression in transfected β cells and that activation was inhibited by a negative regulator of bHLH protein activity, Id. INSAF can also participate in specific protein-protein interactions with the ubiquitous bHLH protein, E12, a factor present in the endogenous β-cell activator factor-ICE DNA complex (11, 18, 42). Using affinity-purified INSAF antibodies, we have also found that the INSAF protein or an antigenically related protein is present in the ICE activator present in β cells. The nuclear localization of INSAF staining in islet α and β cells is also consistent with the cellular distribution of the ICE activator.

The ICE activator appears to be a complex composed of INSAF and the ubiquitous E2A proteins. We expected that the INSAF protein would contain a well-defined bHLH motif, since this region is involved in protein-protein interactions and DNA binding. The HLH domain is composed of two peptide segments capable of forming amphipathic α helices connected by a nonconserved loop region and is required for E2A protein association (29, 30). The basic region is amino terminal to the HLH domain and is necessary for DNA binding. The region between amino acids 214 and 280 of INSAF has some similarity to the bHLH regions of a number of proteins in the bHLH family, such as c-Myc, MyoD, E12, E47, and ITF-1 (Fig. 10) Id, the negative regulator of bHLH proteins, also shows some similarity to INSAF within its HLH domain. However, there are some key amino acid differences between the bHLH region of INSAF and the bHLH proteins listed. The potential basic region of INSAF contains only three conserved residues, and the potential HLH region of INSAF contains only three conserved residues in helix 1 and five conserved residues in helix 2. Presently, we are determining whether this region is important in INSAF-E12 interactions and ICE binding. It is possible that a region distinct from amino acids 214 to 280 of INSAF, such as the potential leucine zipper region identified from amino acids 1 to 61 at the N terminus of INSAF, is involved in E2A protein binding, since the myogenic bHLH activators have been shown to interact through their bHLH region with the c-Jun protein, a factor in the basic leucine zipper family (6, 28). Interestingly, c-Jun also appears to interact with proteins involved in ICE activation (22).

INSAF stimulated ICE-dependent expression only four- to fivefold in HIT T-15 cells (Fig. 3). We are uncertain why the levels of activation were low. It may be that the carboxy-terminal sequences encoded by the missing INSAF 3′ cDNA sequences are important in *trans* activation or that the human INSAF transcription factor does not function effectively in rodent β cell lines and/or that the levels of the transfected INSAF factor are only partially limiting in our assay systems. Another possibility is that INSAF is not the bona fide transcriptional regulator but is antigenically and functionally related to it. Park and Walker (34) have reported that a 25-kDa protein is also a potential heteromeric partner of the E2A protein in a hamster β cell line. INSAF and this protein may represent a family of proteins binding to the ICE. Furthermore, since INSAF activated ICE-mediated expression only in β cells (Fig. 3) and this ICE-mediated expression is negatively regulated in non-insulin-producing cells (55), we considered that a factor present in these cells may prevent activation from the ICE. However, both MyoD and the general bHLH activator, ITF-1, stimulated ICE-mediated expression in non-insulin-producing cell lines (Fig. 4). These results suggest that INSAF activation is specifically blocked in non-insulin-producing cells. The β-cell-specific *trans*-activation pattern of INSAF may indicate that its activity is regulated either through a posttranslational modification(s) and/or by a coregulator.

We detected a uniform distribution for INSAF in the nuclei of cells within the core of the adult mouse islet but not in the acinar cells of the exocrine pancreas. The islet is composed mainly of α and β cells, and we detected nuclear staining for INSAF in both of these cell types. The expression of INSAF in α cells could be related to the developmental history of pancreatic islet cells in vivo. The mouse pancreas develops from the embryonic gut beginning at day 10 of gestation (embryo day 10 [37]). Cells containing glucagon are first seen at embryo day 10, and cells containing insulin also differentiate at this time (1, 37, 49). The other two islet-specific hormones, somatostatin and pancreatic polypeptide, appear in pancreatic

cells at embryo day 16 and birth, respectively (47, 48). During islet cell development, precursor cells express multiple pancreatic hormones (1). The presence of INSAF in an α cell line suggests that this factor is induced in precursor cells at embryo day 10, possibly along with the transcription factors that are required for glucagon expression. The observation that transcription factors that are required for glucagon expression are also present in cloned β cells lines (36) is consistent with this proposal.

The lack of insulin gene expression in α cells appears to be due, at least in part, to the absence of other essential transcriptional activator factor(s); however, it is possible that negative regulatory factors are also important in control (38). Since the function of the ICE appears to be conserved among all mammalian insulin genes, these results indicate that transcription of the insulin gene cannot be induced by the ICE activator alone but instead appears to be mediated by the combined action of this factor and other transcriptional regulators interacting with the insulin gene enhancer. Interestingly, proteins essential for muscle cell differentiation are also present in muscle progenitor cells but in a form unable to stimulate the muscle-specific regulatory cascade (52). The transcriptional activation properties of these muscle-specific regulatory factors appear to be limited in these cells by the expression of dominant-negative regulators of bHLH function, such as Id (5) and c-Jun (6, 28). A future objective will be to define how INSAF-mediated activation is regulated in β cells.

ACKNOWLEDGMENTS

We are grateful to William Taylor and Wayne Anderson for constructive criticism of the manuscript; to George Pendergast for assistance in determining the bHLH region; to Andrea Cherrington, J. P. Johnson, Harold Prow, Howard Masuoka, Jeanne Santoli, and Luibov Lyandvert for expert technical assistance; to Cornelis Murre and Tom Kadesch for anti-E12/E47 antiserum; to Tom Kadesch for the His_6ITF-1 expression clone; and to Alvin Powers for the gift of the human insulinoma library.

This work was supported by grants from the Juvenile Diabetes Foundation and Public Health Service grants GM-30257 (to R.S.), DK-42502 (to R.S. and P.A.W.), and DK-38171 (to G.T.) from the National Institutes of Health. Partial support was also derived from the Vanderbilt University Diabetes Research and Training Center Molecular Biology Core Laboratory (Public Health Service grant P60 DK20593 from the National Institutes of Health).

REFERENCES

1. **Alpert, S., D. Hanahan, and G. Teitelman.** 1988. Hybrid insulin genes reveal a developmental lineage for pancreatic endocrine cells and imply a relationship with neurons. Cell **53**:295–308.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
3. **Andersson, S., D. L. Davis, H. Dahlback, H. Jornvall, and D. W. Russell.** 1989. Cloning, structure and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. J. Biol. Chem. **264**:8222–8229.
4. **Beckmann, H., L.-K. Su, and T. Kadesch.** 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev. **4**:167–179.
5. **Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub.** 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell **61**:49–59.
6. **Bengal, E., L. Ransone, R. Scharfmann, V. J. Dwarki, S. J. Tapscott, H. Weintraub, and I. Verma.** 1992. Functional antagonism between c-*Jun* and MyoD proteins: a direct physical association. Cell **68**:507–519.
7. **Bernard, O., S. Cory, S. Gerondakis, E. Webb, and J. M. Adams.** 1983. Sequence of the murine and cellular *myc* oncogenes and two modes of *myc* transcription resulting from chromosomal translocation in B lymphoid tumours. EMBO J. **2**:2375–2383.
8. **Biedler, J. L., L. Helson, and B. A. Spangler.** 1973. Morphology, growth, tumorgenicity and cytogenetics of human neuroblastoma cells in continuous culture. Cancer Res. **33**:2643–2652.
9. **Bucchini, D., O. Madsen, M.-A. Stinnakre, P. Desbois, P. Lores, E. Monthioux, J. Absil, J.-A. Lepesant, R. Pictet, and J. Jami.** 1986. Pancreatic expression of the human insulin gene in transgenic mice. Proc. Natl. Acad. Sci. USA **83**:2511–2515.
10. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
11. **Cordle, S. R., E. Henderson, H. Masuoka, P. A. Weil, and R. Stein.** 1991. Pancreatic β-cell-type-specific transcription of the insulin gene is mediated by basic helix-loop-helix DNA-binding proteins. Mol. Cell. Biol. **11**:1734–1738.
12. **Cordle, S. R., J. Whelan, E. Henderson, H. Masuoka, P. A. Weil, and R. Stein.** 1991. Insulin gene expression in nonexpressing cells appears to be regulated by multiple distinct negative-acting control elements. Mol. Cell. Biol. **11**:2881–2886.
13. **Crowe, D. T., and M.-J. Tsai.** 1989. Mutagenesis of the rat insulin II 5′-flanking region defines sequences important for expression in HIT cells. Mol. Cell. Biol. **9**:1784–1789.
14. **Davis, R. L., P.-F. Cheng, A. B. Lassar, and H. Weintraub.** 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell **60**:733–746.
15. **Edlund, T., M. D. Walker, P. J. Barr, and W. J. Rutter.** 1985. Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5′ flanking sequences. Science **230**:912–916.
16. **Efrat, S., S. Linde, H. Kofod, D. Spector, M. Delannoy, S. Grant, D. Hanahan, and S. Baekkeskov.** 1988. Beta cell lines derived from transgenic mice expressing a hybrid insulin gene-oncogene. Proc. Natl. Acad. Sci. USA **85**:9037–9041.
17. **Fromont-Racine, M., D. Bucchini, O. Madsen, P. Desbois, S. Linde, S., J. H. Nielsen, C. Saulnier, M.-A. Ripoche, J. Jami, and R. Pictet.** 1990. Effect of 5′-flanking sequence deletions on expression of the human insulin gene in transgenic mice. Mol. Endocrinol. **4**:669–777.
18. **German, M. S., M. A. Blanar, C. Nelson, L. G. Moss, and W. J. Rutter.** 1991. Two related helix-loop-helix proteins participate in separate cell-specific complexes that bind the insulin enhancer. Mol. Endocrinol. **5**:292–299.
19. **German, M. S., L. G. Moss, and W. J. Rutter.** 1990. Regulation of insulin gene expression by glucose and calcium in transfected primary islet cultures. J. Biol. Chem. **265**:22,063–22,066.
20. **Hamaguchi, K., and E. H. Leiter.** 1990. Comparison of cytokine effects on mouse pancreatic α-cell and β-cell lines. Diabetes **39**:415–425.
21. **Hanahan, D.** 1985. Heritable formation of pancreatic β-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. Nature (London) **315**:115–122.
22. **Henderson, E., and R. W. Stein.** 1994. c-Jun inhibits transcriptional activation by the insulin enhancer, and the insulin control element is the target of control. Mol. Cell. Biol. **14**:655–662.
23. **Henthorn, P., M. Kiledijian, and T. Kadesch.** 1990. Two transcription factors that bind to the immunoglobulin enhancer μE5/κE2 motif. Science **247**:467–470.
24. **Hwung, Y.-P., Y.-Z. Gu, and M.-J. Tsai.** 1990. Cooperativity of sequence elements mediates tissue specificity of the rat insulin II gene. Mol. Cell. Biol. **10**:1784–1788.
25. **Karlsson, O., T. Edlund, J. B. Moss, W. J. Rutter, and M. D. Walker.** 1987. A mutational analysis of the insulin gene transcription control region: expression in beta cells is dependent on two related sequences within the enhancer. Proc. Natl. Acad. Sci. USA **84**:8819–8823.
26. **Karlsson, O., M. D. Walker, W. J. Rutter, and T. Edlund.** 1989. Individual protein-binding domains of the insulin gene enhancer positively activate β-cell-specific transcription. Mol. Cell. Biol. **9**:823–827.
27. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **277**:680–685.
28. **Li, L., J.-C. Chambard, M. Karin, and E. N. Olson.** 1992. *Fos* and

Jun repress transcriptional activation by myogenin and MyoD: the amino terminus of *Jun* can mediate repression. Genes Dev. **6**:676–689.

29. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, MyoD, and *myc* proteins. Cell **56**:777–783.
30. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58**:537–544.
31. **Nir, U., M. D. Walker, and W. J. Rutter.** 1986. Regulation of rat insulin I gene expression: evidence for negative regulation in nonpancreatic cells. Proc. Natl. Acad. Sci. USA **83**:3180–3184.
32. **Ohlsson, H., S. Thor, and T. Edlund.** 1991. Novel insulin promoter- and enhancer-binding proteins that discriminate between pancreatic α and β cells. Mol. Endocrinol. **5**:897–904.
33. **Pabo, C., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61**:1053–1095.
34. **Park, C. W., and M. D. Walker.** 1992. Subunit structure of cell-specific E box-binding proteins analysed by quantitation of electrophoretic mobility shift. J. Biol. Chem. **267**:15642–15649.
35. **Peshavaria, M., L. Gamer, E. Henderson, G. Teitelman, C. V. E. Wright, and R. Stein.** 1994. XIHbox 8, an endoderm-specific Xenopus homeodomain protein, is closely related to a mammalian insulin gene transcription factor. Mol. Endocrinol. **8**:806–816.
36. **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.
37. **Pictet, R., and W. J. Rutter.** 1972. Development of the embryonic endocrine pancreas, p. 25–66. *In* D. F. Steiner and M. Frenkel (ed.), Handbook of physiology, section 7. American Physiology Society, Washington, D.C.
38. **Robinson, G. L. W. G., M. Peshavaria, E. Henderson, S.-Y. Shieh, M.-J. Tsai, G. Teietelman, and R. Stein.** 1994. Analysis of transcription regulatory signals of the insulin gene: expression of the *trans*-active factor that stimulates insulin control element mediated expression precedes insulin gene transcription. J. Biol. Chem. **269**:2452–2460.
39. **Selden, F. S., M. J. Skoskiewicz, K. B. Howie, P. S. Russel, and H. M. Goodman.** 1986. Regulation of human insulin gene expression in transgenic mice. Nature (London) **321**:525–528.
40. **Sharma, A., and R. Stein.** 1994. Glucose-induced transcription of the insulin gene is mediated by factors required for β-cell-type-specific expression. Mol. Cell. Biol. **14**:871–879.
41. **Shen, G. H., Z. Zukowska-Grojeck, L. Grundemar, R. Hakanson, and C. Wahlestedt.** 1991. C-terminal neuropeptide Y fragment is a mast cell-dependent vasodepressant agent. Eur. J. Pharmacol. **204**:249–256.
42. **Shieh, S.-Y., and M.-J. Tsai.** 1991. Cell-specific and ubiquitous factors are responsible for the enhancer activity of the rat insulin II gene. J. Biol. Chem. **266**:16708–16714.
43. **Singh, H., J. H. LeBowitz, A. S. Baldwin, Jr., and P. A. Sharp.** 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell **52**:415–423.
44. **Stein, R.** Factors regulating insulin gene transcription. 1993. Trends Endocrinol. Metab. **4**:96–100.
45. **Steiner, D. F., S. J. Chan, J. M. Welsh, and S. C. M. Kwok.** 1985. Structure and evolution of the insulin gene. Annu. Rev. Genet. **19**:463–484.
46. **Studier, W. F., A. H. Rosenberg, and J. J. Dunn.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185**:60–89.
47. **Sumner, S., R. Hakanson, and L. I. Larsson.** 1977. Ontogeny of the rat pancreatic polypeptide (PP) cells. Cell Tissue Res. **178**:303–306.
48. **Teitelman, G., T. H. Joh, and D. J. Reis.** 1981. Linkage of the brain-skin-gut axis: islet cells originate from dopaminergic precursors. Peptides **2**:157–168.
49. **Teitelman, G., and J. K. Lee.** 1987. Cell lineage analysis of pancreatic islet cell development: glucagon and insulin cells arise from catecholaminergic precursors present in the pancreatic duct. Dev. Biol. **121**:454–466.
50. **Thayer, M. J., S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, and H. Weintraub.** 1989. Positive autoregulation of the myogenic determination gene MyoD. Cell **58**:241–248.
51. **Vinson, C. R., K. L. LaMarco, P. F. Johnson, W. H. Landschultz, and S. L. McKnight.** 1988. *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. Genes Dev. **2**:801–806.
52. **Weintraub, H. R., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. B. Lassar.** 1991. The MyoD gene family: nodal point during specification of the muscle cell lineage. Science **251**:761–766.
53. **Westin, G., T. Gerster, M. M. Muller, G. Schaffner, and W. Schaffner.** 1987. OVEC, a versatile system to study transcription in mammalian cells and cell-free extracts. Nucleic Acids Res. **15**:6787–6798.
54. **Whelan, J., S. R. Cordle, E. Henderson, P. A. Weil, and R. Stein.** 1990. Identification of a pancreatic β-cell insulin gene transcription factor that binds to and appears to activate cell-type-specific expression: its possible relationship to other cellular factors that bind to a common insulin gene sequence. Mol. Cell. Biol. **10**:1564–1572.
55. **Whelan, J., D. Poon, P. A. Weil, and R. Stein.** 1989. Pancreatic β-cell-type-specific expression of the rat insulin II gene is controlled by positive and negative transcriptional elements. Mol. Cell. Biol. **9**:3253–3259.
56. **Young, R. A., and R. W. Davis.** 1983. Yeast RNA polymerase II genes: isolation with antibody probes. Science **222**:778–782.