

Molecular Characterization of the Promoter Region of a Neuroendocrine Tumor Marker, IA-1

Qing Li, Abner L. Notkins, and Michael S. Lan¹

From the Oral Infection and Immunity Branch, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20892

Received June 24, 1997

IA-1 is an intronless gene, which encodes a 510 amino acid protein with a zinc-finger DNA-binding motif that is expressed in tumors of neuroendocrine origin. The 5'-upstream region of the IA-1 gene was recently sequenced. In this paper, the regulatory elements and the promoter region of the 5'-upstream region were analyzed by use of a series of deletion mutants (ranging from +26 bp to -2090 bp upstream of the IA-1 gene), which were tested in a pituitary tumor cell line, AtT-20, and HeLa cells by transient transfection assays. These experiments showed that a 506 base pair upstream sequence was sufficient for maximal expression of a reporter gene. Multiple known regulatory elements were found within this region including three E boxes and a clustered Sp-1 site. In addition, Southwestern blot analysis, using a radiolabeled promoter sequence (extending from -108 bp to -66 bp) and nuclear extracts from both neuroendocrine and non-neuroendocrine cell lines, revealed four promoter binding proteins designated PBP1, PBP2, PBP3 and PBP4 with molecular weights of 55 kD, 32 kD, 29 kD, and 27/28 kD, respectively. These studies suggest that several different regulatory elements in the 5'-upstream region of the IA-1 gene and at least four different nuclear proteins may be involved in the cell-specific expression of IA-1. © 1997 Academic Press

A novel insulinoma-associated cDNA, IA-1, was recently isolated from a human insulinoma subtraction library (1). This 2.8 kb cDNA encodes a protein with a five Cys₂-His₂ zinc finger DNA-binding motif at the carboxyl terminus and a putative pro-hormone domain with several dibasic amino acids and an amidation signal sequence (Pro-Gly-Lys-Arg) at the amino terminus. The IA-1 gene also was cloned from a human liver genomic library and found to be a single copy intronless

¹ To whom correspondence should be addressed. Fax: (301) 402-4163.

gene (2). Fluorescence in situ hybridization localized the IA-1 gene to human chromosome 20p11.2. Northern blot analysis revealed that IA-1 was expressed in several human insulinoma tissues and many neuroendocrine tumor cell lines including pheochromocytoma, medullary thyroid carcinoma, insulinoma, pituitary tumor, small cell lung carcinoma and medulloblastoma. A recent clinical study on a panel of 64 human lung cancer cell lines showed that IA-1 mRNA was expressed in 97% (30/31) of small cell lung cancer cell lines and 13% (4/30) of non-small cell lung cancer cell lines which were shown to have neuroendocrine features (3).

The fact that IA-1 was expressed primarily in tumors of neuroendocrine origin, suggested that IA-1 might be under tight transcriptional control by cell-specific regulatory elements. Sequence analysis of 2090 bp upstream from the IA-1 transcription start site failed to reveal typical TATA or CCAAT boxes, but did reveal several Sp-1 clusters, multiple T antigen binding sites and other putative neuroendocrine-specific regulatory elements. In this study, a series of deletion mutants ranging from +26 bp to -2090 bp upstream of the IA-1 gene were constructed and tested in transient transfection assays. Nuclear extracts isolated from pituitary tumor and HeLa cell lines were examined by a gel mobility shift assay and cell-specific nuclear factors were identified by Southwestern blot analysis.

MATERIALS AND METHODS

Plasmid constructs. A 13 kb genomic clone containing the IA-1 gene and the flanking sequences were isolated from a human λ GEM-11 liver genomic library (2). A Sac I fragment between -1253 bp and +649 bp was subcloned into a pBlueScript II vector (Stratagene, La Jolla, CA). Using Exo-Mung Bean deletion method, the 3' end of this clone was deleted to +169 bp that is 7 bp downstream of the translation start site (+162 bp). The modified construct was designated as IA1-1253-E that contains IA-1 gene 5'-upstream sequence from -1253 bp to +169 bp. A series of progressive 5' end deletions of IA1-1253-E were prepared by Exo-Mung Bean nest deletion method (4, 5). To prepare the deletion mutants, IA1-1253-E plasmid was digested by restriction en-

zymes, Not I and Pst I. The Not I digestion produced a 5'-overhangs and further deleted by Exo nuclease III treatment whereas the Pst I digestion produced a 3'-overhangs that was not recognized by Exo nuclease III and thus protect the vector from being further deleted. Deletion was performed by Exo nuclease III digestion for various times at 4° C. The reaction mixture was followed by Mung Bean treatment to blunt the both ends. The nest deleted plasmid was re-ligated and transformed into XL-1-Blue competent cells. Deletion clones were characterized by DNA sequencing.

The deletion mutants were subsequently ligated to a promoterless luciferase reporter vector, pGL2-E plasmid (Promega, Madison, WI). A recombinant plasmid that contains fragment between -2090 bp and +169 bp was constructed by ligating two Sac I fragments isolated from IA-1 genomic clone, which contains sequences from -2090 bp to -1253 bp and from -1253 bp to +169 bp, respectively. The orientation of the sequence was determined by sequencing at the junction, and the fragment with correct orientation was then ligated to pGL2-E plasmid, designated as IA1-2090-E. Totally, ten deletion mutants were chosen to perform the transient transfection assay.

Cell culture and transfection assay. Cell lines, AtT-20 and GH3 were obtained from the American Type Culture Collection (Rockville, MD); HeLa and NIH3T3 were obtained from Dr. R. S. Metzgar (Duke University, NC). A mouse insulinoma cell line, β TC-1, was kindly provided by Dr. E. H. Leiter (Bar Harbor, ME). Rat insulinoma cell line, RIN (6), was propagated routinely in our laboratory. Cell lines were cultured in modified Eagle's medium with penicillin G (100 units/ml) and streptomycin sulfate (100 μ g/ml). The medium for AtT-20 and GH3 was supplemented with 10% horse serum and the medium for other cell lines was supplemented with 10% fetal bovine serum.

For transient transfection assay, cells were plated on 6 cm plate 24 hours prior to transfection at approximately 50-70% confluence. Cells were washed twice with DMEM medium without serum. Fourteen microliters of transfectam reagent (Promega), 4 μ g tested DNA and 3 μ g pSV- β -galactosidase DNA as an internal control were suspended in DMEM medium without serum and then added to each plate with final volume of 1.5 ml. Transfection was performed from 4 hours for HeLa cells to 10 hours for AtT-20 cells. The transfection was stopped by adding 5 ml of culture medium with serum and incubated for additional 24 hours. Cells were washed twice with PBS and lysed with 0.25ml 1 \times Reporter buffer (Promega) at room temperature for 15 minutes. The plates were scraped and cell lysate were collected and centrifuged at 12,000 g for 2 minutes. The supernatants were collected for the luciferase assay and the β -galactosidase assay as described (Promega). Transfection efficiency was normalized by internal standard, β -galactosidase activity.

Gel mobility shift assay. Nuclear extracts from culture cell lines were prepared as described (7). Briefly, culture cells were scraped from flask and washed twice in phosphate-buffered saline. The cells were resuspended in lysis buffer (10 mM Tris, pH 7.6, containing 10 mM NaCl, 3 mM MgCl₂, 0.5% NP40, 0.3 mM DTT and 0.5 mM PMSF) for 5 minutes on ice and placed in a dounce homogenizer for 30 strokes. The suspension was centrifuged at 4000 rpm for 15 minutes and the pellets resuspended in extraction buffer (20 mM HEPES pH 7.9, containing 20% glycerol, 0.4 M NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM DTT and 0.5 mM PMSF) on ice for 1 hour. The solution was centrifuged at 14,000 rpm for 30 minutes and the supernatant was collected for dialysis in 20 mM HEPES (pH 7.9, containing 20% glycerol, 50 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT and 0.5 mM PMSF). The concentration of the nuclear extracts was determined by BCA protein assay (Pierce, Rockford, IL) and aliquots were stored in the -70° C freezer. Two complementary oligonucleotides from -108 bp to -66 bp and from -216 bp to -253 bp were synthesized by Bio-synthesis Inc. (Lewisville, TX) and the annealed probes (PE-5 and PE-8) were labeled with ³²P-dCTP using Klenow fragment (Life Technologies, Inc.,

Gaithersburg, MD) or Stoffel fragment (Perkin-Elmer Corp., Norwalk, CT) by filling in 5' protruding ends. Ten microgram proteins of each nuclear extract with or without the addition of anti-Sp-1 antibody were incubated with radiolabeled probe (20,000 cpm) for 20 minutes at room temperature, the protein-DNA complex then was applied to a 4% non-denaturing polyacrylamide DNA gel containing 2.5% glycerol and buffered with 0.5 \times TBE (45 mM Tris base, 45 mM boric acid, 1 mM EDTA). The electrophoresis was carried out at 4° C for 3 hours. For competition experiments, excess amount of non-radiolabeled probe was incubated with the protein for 15 minutes prior to the addition of the ³²P-labeled probe at room temperature for 15 minutes. The gel was dried on the filter paper and exposed overnight to the film at -70° C. The following double-stranded sequences were used as probes in the gel mobility shift competition assay and Southwestern blot analysis:

PE-5 (-108 bp to -66 bp, IA-1 promoter core sequence):
5'-AGGCTGAGGAGCTGCGGACGCGCTGATTGGCTCCAG-GGAAGC-3'

PE-8 (-216 bp to -253 bp, Sp1 clusters):
5'-TGGACCGGGCGGGGCGGCGCGCCGGGCGGGGCGAC-CA-3'

Sp-1: 5'-ATTCGATCGGGGCGGGGCGAGC-3'

Southwestern blot analysis. PE-5, end-labeled with ³²P-dCTP, served as the probe. Nuclear extracts were prepared as described above. Nuclear extracts (30 μ g) were electrophoresed at 125 volts in a 10% SDS-PAGE gel at room temperature for 1.5 hours and then electrotransferred to a nitrocellulose membrane at 40 volts for 5 hours at 4° C. The blot was blocked overnight in 5% non-fat dry milk in 10 mM HEPES buffer (pH 8.0), and then incubated with binding buffer (10 mM HEPES buffer, pH 8.0, containing 50 mM NaCl, 10 mM MgCl₂, 0.1 mM EDTA, 1 mM DTT and 0.25% non-fat dry milk). Radiolabeled probe was added at a concentration of 5.0 \times 10⁶ cpm/ml into the binding buffer at 4° C for 4 hours. The filter was washed with 10 mM HEPES, pH 8.0, containing 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT and 0.25% non-fat dry milk twice, then exposed to Kodak X-OMAT film overnight.

RESULTS

Mapping the Functional Regulatory Elements of the IA-1 Gene

To identify sequences at the 5'-flanking region that control transcription of the IA-1 gene, a series of 5' deletion mutants were constructed. Deletion mutants ranging from -2090 bp to +26 bp at the 5' end and shared a common 3' end at +169 bp were fused to a promoterless vector (pGL2-E) containing firefly luciferase reporter gene and SV40 enhancer (Figure 1A). Reporter constructs and pSV β -gal internal control were introduced into AtT-20 cells, a pituitary cell line which expresses the IA-1 gene, and HeLa cells which do not express the IA-1 gene. Transfection were repeated four to six times and each assay was normalized according to the internal control, pSV β -gal, for transfection efficiency. Plasmid, pGL2-C, containing SV40 promoter and enhancer was used as a positive control in each assay. Sequence -506 bp was sufficient for maximal reporter gene expression in AtT-20 cells (Figure 1B). All the constructs longer than -506 bp demonstrated higher activity than the positive control vector that contained both SV40 promoter and enhancer. Deletion of the sequence be-

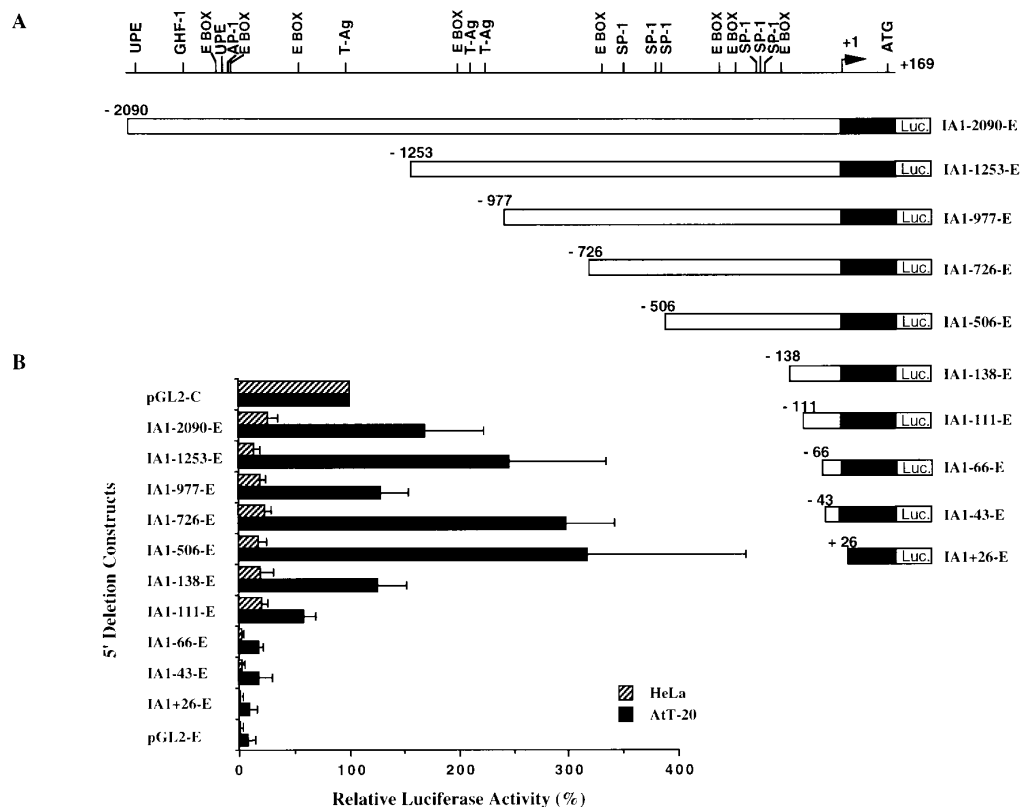


FIG. 1. IA-1 gene 5'-upstream deletion mutants and their relative promoter activities in AtT-20 and HeLa cells. (A) IA-1 gene 5'-flanking region is nest deleted from -2090 bp to $+26$ bp. The various deleted fragments that contain the same 3' end at $+169$ bp were placed upstream of the luciferase structure gene with a sense orientation and subjected to transient transfection assay. The putative regulatory elements are labeled at the top of the diagram, the transcription initiation site (+1) is marked by an arrow and the translation start codon indicated by ATG. (B) Transfection analysis of the IA-1 deletion mutants in neuroendocrine (AtT-20) and non-neuroendocrine (HeLa) cells. The reporter gene constructs were introduced into AtT-20 and HeLa cells in parallel with pGL2-C which contains the SV40 promoter and enhancer for transient transfection assay. The pSV β -gal plasmid was co-transfected along with each test DNA as an internal control. The luciferase activity was normalized to β -galactosidase activity. The results were expressed as percentage of the luciferase activity compared with the positive control vector, pGL2-C. The data for each construct represents the average of at least four different experiments.

tween -977 bp to -726 bp increased the reporter expression by 120% relative to expression directed by construct IA1-977-E. A possible negative regulatory element may reside in that region. Deletion to -138 bp resulted in a 60% reduction in the maximal activity indicating the existence of the positive regulatory elements. Sequence analysis revealed three clustered Sp-1 binding sites (GC box) in the region.

When we further deleted the sequence from -138 bp to -111 bp, another 50% reduction of transcriptional activity was seen. A three fold reduction in activity was shown when sequences between -111 bp and -66 bp were removed. Plasmid, IA1+26-E, containing $+26$ bp to $+169$ bp, failed to direct reporter gene expression as compared to the negative control vector, pGL2-E. The mapping data showed that the maximal activity of the reporter gene occurred when the flanking sequence up to 506 bp was included. The gradual loss of transcrip-

tional activity between -506 bp and -66 bp suggests that other positive regulatory elements exist in addition to the Sp-1 cluster.

To determine whether the Sp-1 protein binds to the Sp-1 cluster, gel shift experiments were performed with the addition of competitive Sp-1 oligonucleotide or anti-Sp-1 antibody. The sequence from -216 bp to -253 bp, containing the three GC boxes, was synthesized, radiolabeled and used as a probe (PE-8). As seen in Figure 2, nuclear extract prepared from both HeLa and AtT-20 cells formed specific shift bands (lane 2 and 6) that were competed out by either unlabeled Sp-1 oligonucleotide (lane 4 and 8) or unlabeled PE-8 probe (lane 5 and 9). When the anti-Sp-1 antibody was added to the reaction mixture, a supershifted band was formed by an antibody-protein-DNA complex (lane 3 and 7). The fact that Sp-1 antibody supershifted the protein-DNA complex suggests

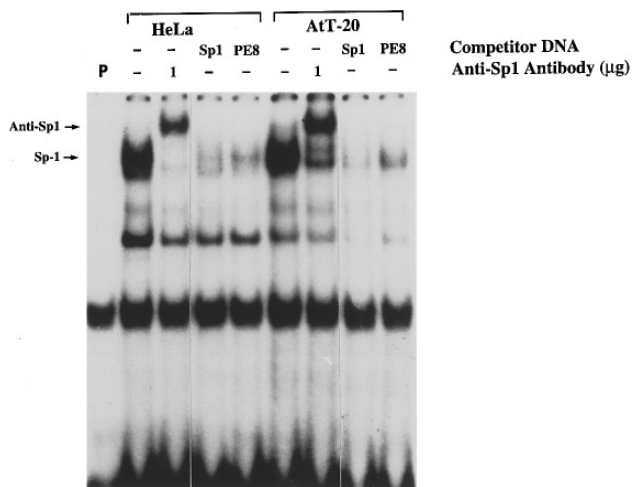


FIG. 2. Sp-1 protein binds to the GC boxes of the sequence between -216 bp and -253 bp. Complementary oligonucleotides representing the sequence from -216 bp to -253 bp of the IA-1 gene was annealed and labeled as a probe, PE-8. Ten microgram of nuclear extract isolated from HeLa or AtT-20 cells was mixed with 20,000 cpm PE-8 probe with or without anti-Sp-1 antibody (1 µg) or 90-fold molar excess of competitor DNA. Shifted and supershifted bands are indicated by arrows. The first lane represents probe only as indicate by "P".

that the major protein that binds to the IA-1 PE-8 probe is the Sp-1 protein.

Binding of IA-1 Promoter to the Nuclear Extracts of Neuroendocrine and Non-neuroendocrine Cell Lines

To look for proteins that might bind to other regions of the IA-1 promoter, the sequence located between -108 bp and -66 bp was used and designated PE-5. Gel shift analysis showed a major DNA protein complex designated A that formed with both HeLa and AtT-20 nuclear extracts (Figure 3). Cold probe, PE-5, competed to inhibit the formation of this complex. A minor complex migrating immediately below complex A was seen with nuclear extracts isolated from AtT-20 but not HeLa cells, when higher concentrations of the extracts (10 and 15 µg) were used. These findings suggest that some unique nuclear proteins from AtT-20 cells bind to the IA-1 promoter region.

Detecting the Nuclear Proteins That Bind Specifically to the PE-5 Region of the IA-1 Promoter

Although the gel mobility shift assay indicated specific binding of nuclear proteins to the PE-5 sequence, it remains unclear how many proteins are involved in forming the DNA-protein complexes. To address this question, Southwestern blot analysis was carried out to identify nuclear proteins that recognize the IA-1 promoter sequence, PE-5. Nuclear extracts isolated from IA-1 expressing cell lines (AtT-20, GH3, βTC-1 and

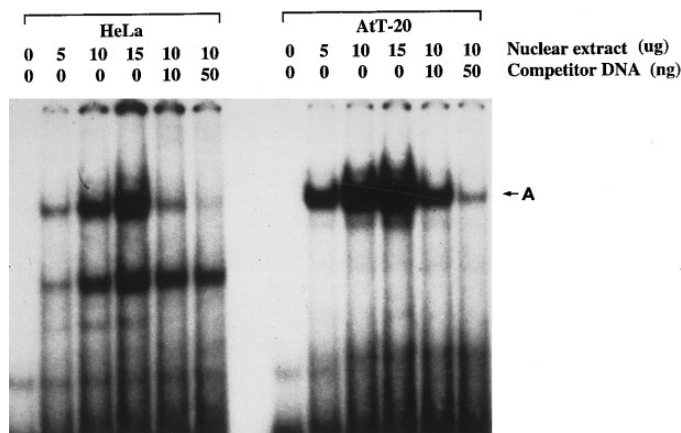


FIG. 3. Gel mobility shift assay of AtT-20 and HeLa cell nuclear extracts with IA-1 promoter. The radiolabeled oligonucleotide probe, PE-5, representing IA-1 promoter sequence from -108 bp to -66 bp, was mixed with various concentration of nuclear extracts and cold PE-5 probe (competitor DNA). A major DNA-protein complex, designated A, was observed.

RIN cells) and IA-1 non-expressing cell lines (HeLa and NIH3T3 cells) were analyzed with radiolabeled PE-5 probe (Figure 4). An IA-1 promoter binding nuclear protein, designated PBP4, with a molecular weight of approximately 27/28 kD, was found in neuroendocrine tumor cell extracts, but not in extracts of HeLa and

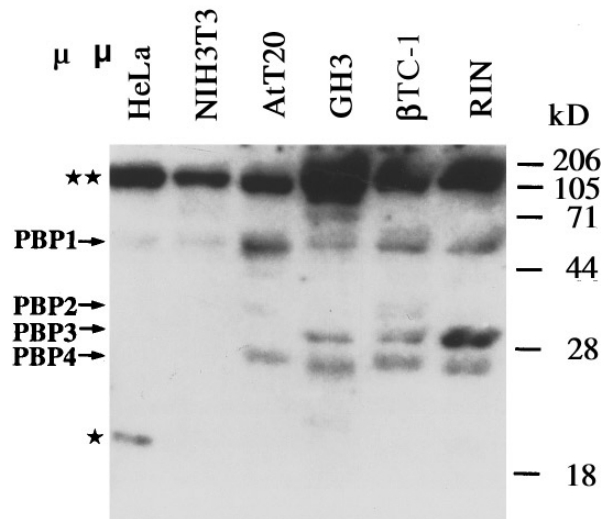


FIG. 4. Southwestern blot analysis of nuclear proteins that interacted with the IA-1 promoter. Nuclear proteins (30 µg) isolated from HeLa, NIH3T3, AtT-20, GH3, βTC-1 and RIN cells were electrophoresed in a SDS-PAGE gel and electrotransferred to a nitrocellular membrane. Probe, PE-5, was end-labeled by filled-in reaction. Proteins recognized by PE-5 probe were designated as PBP1, PBP2, PBP3 and PBP4. "★" indicates a low molecular weight protein specific for HeLa cells. A high molecular weight protein (★★) was presented in all the cell lines tested.

NIH3T3 cells. A 29 kD protein, designated PBP3, was found in three of the four IA-1 expressing cells (i.e., GH3, β TC-1 and RIN), but was absent in AtT-20 and IA-1 non-expressing cells. A 32 kD nuclear protein, designated PBP2, was found in nuclear extracts of AtT-20 and β TC-1, both cells of mouse origin. A 55 kD non-cell-type-specific protein, designated PBP1, was found in all the cell line extracts tested in the present experiments. The intensity of the signal, however, was significantly stronger in IA-1 expressing cells than in IA-1 non-expressing cells. An even higher molecular weight protein (over 105 kD) that bound to PB5 was found in all the cell lines tested. These data show that the PE-5 probe recognizes three different proteins, PBP2, PBP3 and PBP4, that appear to be neuroendocrine cell-specific.

DISCUSSION

IA-1 gene expression appears to be restricted to tumors of neuroendocrine origin. Recently, we sequenced the 5'-upstream sequence of IA-1 and showed by comparison with known sequences in the gene bank that it contains multiple regulatory elements that could contribute to tissue-specific expression (2). The current experiments, employing transient transfection with the 5'-upstream region linked to a reporter gene, revealed that three regions, -1253 bp to -977 bp, -978 bp to -726 bp and -506 bp to -111 bp, contain cis-acting regulatory elements. The sequence between -506 bp and -111 bp containing multiple E boxes, SP-1 sites, and proximal sequence adjacent to the transcriptional start site, are required for maximal expression of the reporter gene.

Sequence analysis of this region revealed three E boxes (-177 bp to -182 bp, -310 bp to -315 bp and -356 bp to -361 bp) and a clustered Sp-1 site (-216 bp to -253 bp). E boxes contain the CANNTG motif which is the binding site motif for proteins in the basic helix-loop-helix (b-HLH) family (8). E boxes function as both positive and negative elements (9-11) and are involved in the response to glucose stimulation (12). The most proximal E box in the IA-1 gene 5'-upstream sequence contains a CATCTG motif that is perfectly matched to the E box of insulin gene, also named IEB box (insulin enhancer binding box). The IEB box has been shown to be the most important enhancer element for basal and tissue specific insulin gene transcription (13, 14, 11). Since IA-1 gene expression is restricted to insulinoma and other neuroendocrine cells, it is likely that the E boxes found between -506 bp and -138 bp play a role in enhancing transcription and tissue specificity.

Clustered Sp-1 sites are found in the promoter region of many TATA-less genes (15-17). Sp-1 protein is ubiquitously distributed and a clustered Sp-1 site is known to be sufficient to transactivate the insulin-like growth

factor binding protein-2 promoter (18). Since the IA-1 gene is a TATA-less gene and the clustered Sp-1 sites are located proximal to the IA-1 transcriptional start site, we performed a gel shift experiment to determine the potential binding activity of Sp-1 sites. Using both the Sp-1 cluster sequence and the commercial Sp-1 oligonucleotide, we demonstrated that the protein bound to the Sp-1 cluster site was the Sp-1 protein and could be supershifted by anti-Sp-1 antibody. This argues that the Sp-1 protein may contribute to the enhancer activity of IA-1 gene expression.

In addition, we showed that the region immediately before the transcription initiation site (-108 bp to -66 bp) possesses promoter activity. Moreover, by comparing nuclear extracts from pituitary and HeLa cells by gel mobility shift analysis, we found that the expression of the IA-1 gene is cell type-dependent. This suggests that different regulatory proteins may be involved in the expression of IA-1 in different neuroendocrine cell lines. To further identify the regulatory proteins that interact with the IA-1 PE-5 sequence, Southwestern assays were carried out with nuclear proteins from pituitary, insulinoma, HeLa and NIH3T3 cells. At least three nuclear proteins, PBP2, PBP3 and PBP4 with molecular masses of 27/28 kD, 29 kD and 32 kD, respectively, appear to be involved in the expression of the IA-1 gene in pituitary and insulinoma cells. Currently, we are attempting to characterize these nuclear proteins.

ACKNOWLEDGMENTS

We thank Drs. P. Zhou and B. W. Zhang for critical reading of the manuscript. We gratefully acknowledge the editorial help of Janice Solomon.

REFERENCES

- Goto, Y., Desilva, M. G., Toscani, A., Prabhakar, B. S., Notkins, A. L., and Lan, M. S. (1992) *J. Biol. Chem.* **267**, 15252-15257.
- Lan, M. S., Li, Q., Lu, J., Modi, W. S., and Notkins, A. L. (1994) *J. Biol. Chem.* **269**, 14170-14174.
- Lan, M. S., Russell, E. K., Lu, J., Johnson, B. E., and Notkins, A. L. (1993) *Cancer Res.* **53**, 4169-4171.
- Laskowski, M., SR. (1980) *Methods Enzymol.* **65**, 263-276.
- Rogers, S. G., and Weiss, B. (1980) *Methods Enzymol.* **65**, 201-211.
- Gazdar, A. F., Chick, W. L., Qie, H. K., Sims, H. L., King, G., Weir, G. C., and Lauris, V. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3519-3523.
- Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 11475-1489.
- Murre, C., McCaw, P. S., and Baltimore, D. (1989) *Cell* **56**, 777-783.
- Hwang, Y. P., Gu, Y. Z., and Tsai, M. J. (1990) *Mol. Cell. Biol.* **10**, 1784-1788.
- Karlsson, O., Walker, M. D., Butter, W. J., and Edlund, T. (1989) *Mol. Cell. Biol.* **9**, 823-827.

11. Whelan, J., Poon, D., Weil, P. A., and Stein, R. (1989) *Mol. Cell Biol.* **9**, 3253–3259.
12. German, M. S., and Wang, J. (1994) *Mol. Cell Biol.* **14**, 4067–4075.
13. Crowe, D. T., and Tsai, M. J. (1989) *Mol. Cell Biol.* **9**, 1784–1789.
14. Karlsson, O., Edlund, T., Moss, J. B., Rutter, W. J., and Walker, M. D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8819–8823.
15. Blake, M. C., Jambou, R. C., Swick, A. G., Kahn, J. W., and Azizkhan, J. C. (1990) *Mol. Cell Biol.* **10**, 6632–6641.
16. Chen, X., Azizkhan, J. C., and Lee, D. C. (1992) *Oncogene* **7**, 1805–1815.
17. Lee, J. K., Tam, J. W. O., Tsai, M. J., and Tsai, S. Y. (1992) *J. Biol. Chem.* **267**, 4638–4645.
18. Boisclair, Y. R., Brown, A. L., Casola, S., and Rechler, M. M. (1993) *J. Biol. Chem.* **268**, 24892–24901.