

Genomic Organization, 5'-Upstream Sequence, and Chromosomal Localization of an Insulinoma-associated Intronless Gene, IA-1*

(Received for publication, December 8, 1993)

Michael S. Lan[‡], Qing Li, Jia Lu, William S. Modi[§], and Abner L. Notkins

From the Laboratory of Oral Medicine, NIDR, National Institutes of Health, Bethesda, Maryland 20892 and the [§]Biological Carcinogenesis Development Program, Program Resources, Inc./DynCorp, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201

IA-1 is a novel cDNA originally isolated from a human insulinoma subtraction library (ISL-153). It encodes a protein containing both a zinc finger DNA-binding domain and a putative prohormone domain. IA-1 transcripts have been found thus far only in tumors of neuroendocrine origin. Clinical studies have shown that IA-1 is a sensitive marker for neuroendocrine differentiation of human lung tumors.

In this study, we cloned and sequenced the entire IA-1 gene and its 5'-upstream region from a human liver genomic library. *In situ* hybridization localized the IA-1 gene to the short arm of human chromosome 20. Sequence analysis and restriction enzyme mapping showed that the IA-1 gene is uninterrupted and appears to be intronless. Evidence that IA-1 is an intronless gene that can translate into protein was obtained from *in vitro* translation studies that showed that both IA-1 cDNA and IA-1 genomic DNA yielded identical protein products of approximately 61,000 daltons. Examination of the 5'-upstream region (2090 base pairs) revealed several tissue-specific regulatory elements, including glucocorticoid upstream promoter elements and a Pit-1 factor binding site. The presence of several different upstream regulatory elements may account for IA-1 gene expression in different neuroendocrine tumors.

We have recently reported the sequence of a novel human insulinoma-associated cDNA, IA-1,¹ that encodes a protein with zinc finger DNA-binding motifs (Goto *et al.*, 1992). This unique cDNA has revealed several interesting features, which include a putative prohormone domain with several dibasic amino acids, an amidation signal sequence (Pro-Gly-Lys-Arg) at the amino terminus, and five Cys₂-His₂ zinc finger DNA-binding motifs at the carboxyl terminus. Expression of the IA-1 gene was detected by Northern analysis in five of five human insulinoma tissues, in all of the murine insulinoma cell lines (including mouse, rat, and hamster), and in other neuroendocrine tumor cell lines (such as pituitary tumor, pheochromocytoma, medullary thyroid carcinoma, and small cell lung carcinoma), but it was not detected in a variety of other tumors (such as melanoma, breast carcinoma, thyroid carcinoma, glioblastoma, pancreatic carcinoma, and choriocarcinoma) or in

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) U07172.

[‡] To whom correspondence should be addressed: Laboratory of Oral Medicine, NIDR, NIH, Bldg. 30, Rm. 121, Bethesda, MD 20892.

¹ The abbreviations used are: IA-1, insulinoma-associated cDNA-1; bp, base pair(s); kb, kilobase(s); PIPES, 1,4-piperazinediethanesulfonic acid.

normal tissues (including pancreas, testes, lymph node, brain, lung, liver, stomach, spleen, thyroid, kidney, and colon). Clinical studies on a panel of 64 human lung cancer cell lines by Northern analysis demonstrated that IA-1 mRNA was detected in 97% (30/31) of small cell lung cancer cell lines and 13% (4/30) of non-small cell lung cancer cell lines (Lan *et al.*, 1993). The IA-1-positive non-small cell lung cancer cell lines were shown to have neuroendocrine features as evidenced by the expression of one or more known neuroendocrine markers, such as chromogranin A and L-dopa decarboxylase. The restricted expression of the IA-1 gene in neuroendocrine tumors suggests that IA-1 gene expression may be under the tight control of certain tissue-specific regulatory elements.

In this study, we cloned and characterized the IA-1 gene from a human liver genomic library. We mapped, by restriction enzymes, a 13-kb genomic fragment that contained the IA-1 gene, sequenced the entire gene and its 5'-upstream region, and localized the IA-1 gene to chromosome 20 p11.2 by fluorescence *in situ* hybridization.

MATERIALS AND METHODS

Screening of IA-1 Genomic Clones from Normal Human Liver Library—A human liver genomic library in λ GEM-11 was purchased from Promega (Madison, WI). This library contains 9–23-kb fragments of normal human genomic DNA partially digested with *Sau3AI* and two bases filled in by Klenow fragment. The total primary plaque-forming units of this library is 2.0×10^6 , and the final amplified titer is 1.65×10^9 plaque-forming units/ml. The library was screened by the standard plaque hybridization method (Sambrook *et al.*, 1989). Briefly, duplicate filters were lifted, baked under vacuum at 80 °C for 1 h, and washed with $5 \times$ SSC for 30 min. The filter was prehybridized at 50 °C with 40% formamide, $5 \times$ SSC, 10 μ g/ml sheared salmon sperm DNA, 6 \times Denhardt's solution (5 Prime \rightarrow 3 Prime, Inc., Boulder, CO) for 6 h followed by hybridization in the same solution with a randomly primed labeled (10^6 cpm/ml) cDNA probe, IA-1-34 (Goto *et al.*, 1992). Hybridization was carried out at 50 °C for 16 h, and the filter was washed with $1 \times$ SSC, 0.1% SDS at the same temperature. Positive plaques from duplicate filters were subjected to secondary and tertiary screening.

DNA Sequence Analysis and Restriction Enzyme Mapping—A positive genomic clone, containing a 13-kb insert, was digested with *SacI* and subcloned into the pBlueScriptII (sk+) vector (Stratagene, La Jolla, CA). Plasmid DNA was used for double strand DNA sequencing. DNA sequencing was performed using Sequenase T4 DNA polymerase under conditions recommended by the supplier (U. S. Biochemical Corp.). Internal sense and antisense strand primers were synthesized by Biosynthesis, Inc. (Lewisville, TX). Both strands were sequenced and analyzed using the GeneWorks 2.1 software package (IntelliGenetics, Mountain View, CA). The current FASTA data base was used for searching both nucleic acid and protein sequence similarities (Pearson and Lipman, 1988). Restriction enzyme mapping of the isolated genomic clone was performed by the method of Whittaker and Southern (1986) with some modifications. Briefly, plasmid DNA from the pL3S5.6 or pL3S4.9 clone was dissolved in 100 μ l of $1 \times$ restriction buffer containing 10 mM magnesium chloride. Twenty- μ l aliquots were UV cross-linked for 0, 1, 15, 45, or 60 min using Stratalinker (Stratagene) and subjected to appropriate enzyme digestion. The sample was heated at 65 °C for 15 min and run on a 1% agarose gel for Southern analysis.

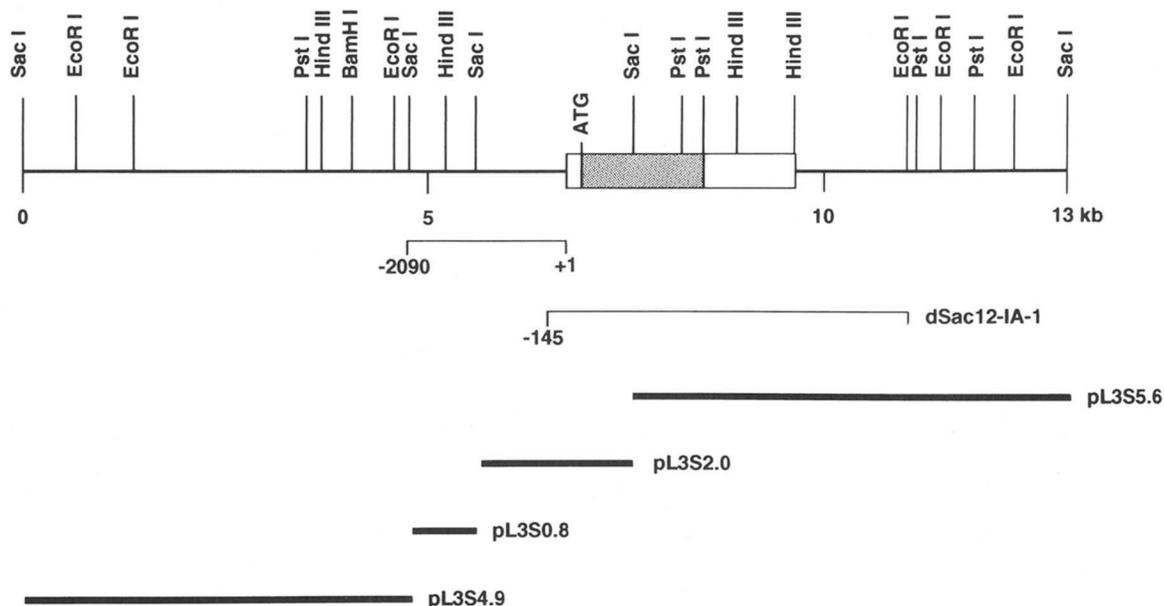


FIG. 1. **Genomic organization of the IA-1 gene.** The restriction enzyme map of a 13-kb genomic clone containing the IA-1 gene was determined by a combination of sequencing and restriction enzyme mapping with *SacI*, *EcoRI*, *PstI*, *HindIII*, and *BamHI* as described under "Materials and Methods." The 13-kb fragment was digested with restriction enzyme *SacI* and subcloned into the pBlueScriptII vector. Four clones, designated pL3S4.9, pL3S0.8, pL3S2.0, and pL3S5.6, were generated. The rectangular box shows the location of IA-1 cDNA; the shaded area represents the coding sequence, and the open areas represent the 5'- and 3'-untranslated regions. The 5'-upstream sequence (from +1 to -2090 bp) of the IA-1 gene is shown in Fig. 6. A clone, *dSac12-IA-1*, was prepared as a DNA template for *in vitro* translation study. No intron was found in the IA-1 gene.

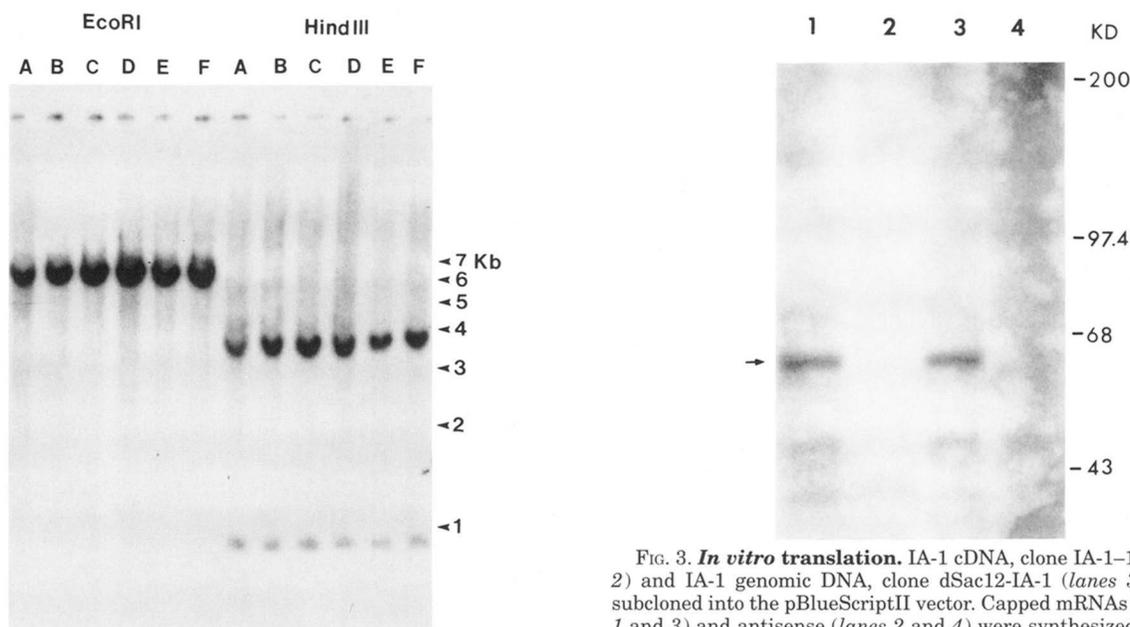


FIG. 2. **Southern analysis.** Fifteen μg of genomic DNA isolated from 16 individuals were digested with restriction enzymes *EcoRI* or *HindIII* and electrophoresed on a 0.8% agarose gel. Denatured DNAs were transblotted to Nytran and hybridized with ^{32}P -labeled full-length IA-1 cDNA. The blot was washed under high stringency and exposed at -70°C for 3 days. Six samples (A-F) are shown.

Hybridization was performed at 39°C for 18 h in a solution containing $6 \times$ Denhardt's solution, $10 \mu\text{g/ml}$ sheared salmon sperm DNA, $2 \times$ SSC, and 10^6 cpm/ml end-labeled probe. Four oligonucleotide probes (TGC-CCACTAGTAACTAAC, AGTGGGAAAAGGATTCCG, TGGGGGACT-CATCTTATC, AAGAAGCCCAAGGCCAT), located at both ends of pL3S5.6 and pL3S4.9, were end-labeled by T4 polynucleotide kinase (Life Technologies, Inc.) and used to probe partially digested DNAs.

Southern Analysis—Total genomic DNAs were prepared from 16 individuals by standard methods (Sambrook *et al.*, 1989). Fifteen μg of DNA were digested with restriction enzymes *EcoRI* or *HindIII* and

FIG. 3. ***In vitro* translation.** IA-1 cDNA, clone IA-1-18 (lanes 1 and 2) and IA-1 genomic DNA, clone *dSac12-IA-1* (lanes 3 and 4) were subcloned into the pBlueScriptII vector. Capped mRNAs of sense (lanes 1 and 3) and antisense (lanes 2 and 4) were synthesized in an *in vitro* transcription system. One μg of each transcript was translated in a rabbit reticulocyte lysate with ^{35}S -cysteine. Translated products were separated on a 7.5% SDS-polyacrylamide gel electrophoresis gel, fixed with autoradiography enhancer, and the film was exposed overnight. The translated protein is marked by an arrow along with protein markers.

electrophoresed on a 0.8% agarose gel. Denatured DNAs were transblotted to Nytran via capillary blotting and hybridized with a ^{32}P -labeled cDNA IA-1-18 probe (Goto *et al.*, 1992). Hybridization was performed at 50°C for 18 h in a solution containing 40% formamide, $5 \times$ SSC, $10 \mu\text{g/ml}$ sheared salmon sperm DNA, $6 \times$ Denhardt's solution, and 10^6 cpm/ml radiolabeled probe. The blot was washed twice in $0.1 \times$ SSC, 0.1% SDS at 60°C (high stringency). Autoradiograms were exposed at -70°C for 3 days.

RNAse Protection Assay—A genomic subclone (*dSac3BE*), including both IA-1 cDNA and 5'-flanking regions, was prepared from *BglI* digestion and 5'-deletion of pL3S2.0. ^{32}P -Labeled antisense riboprobe (ap-

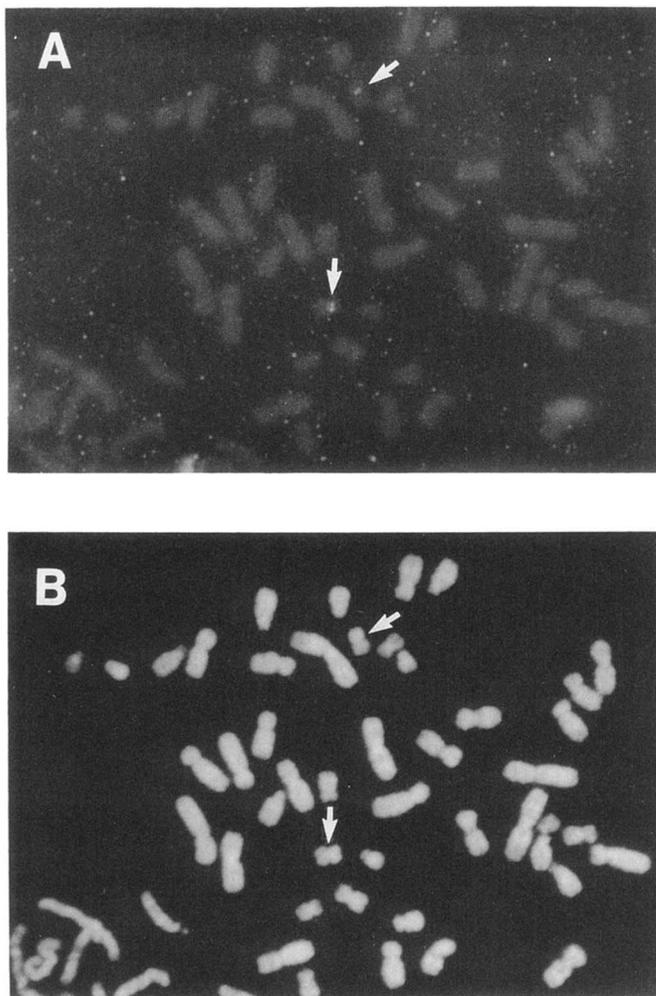


FIG. 4. **Chromosomal localization.** A, metaphase chromosomes following fluorescence *in situ* hybridization with the genomic IA-1 probe. Arrows indicate the signal at 20p11.2. B, simultaneous Hoechst 33258 staining and QFJ-banded chromosomes.

proximately 400 nucleotides) was prepared by T7 RNA polymerase transcription and subjected to 6% urea gel purification. The riboprobe (1×10^5 cpm) was mixed with 50 μ g of HeLa RNA or NCI-H69 small cell lung carcinoma RNA in a hybridization buffer (40 mM PIPES, pH 6.4, 50% formamide). The reaction mixture was heated to 90 °C for 5 min and slowly cooled down to 45 °C for overnight incubation. RNase digestion was carried out by adding 300 μ l of RNase solution (300 mM NaCl, 10 mM Tris-Cl, pH 7.4, 5 mM EDTA, 2 μ g/ml RNase T1, 40 μ g/ml RNase A) at room temperature for 10 min. The reaction was stopped by adding 20 μ l of 1% SDS and 5 μ l of 10 mg/ml protease K at 37 °C for 15 min. The reaction mixture was extracted with 400 μ l of phenol/chloroform twice and ethanol precipitated. The protected fragment was analyzed on an 8% denaturing sequencing gel.

In Vitro Translation.—An *Eco*RI fragment (6.5 kb) from genomic IA-1 DNA was subcloned into the pBlueScriptII (sk+) vector (Stratagene). Serial deletion was performed to eliminate the 5'-untranslated sequence. One clone (designated dSac12-IA-1, starting at position -145 bp and extending to the *Eco*RI site in the 3'-region of IA-1 gene) was chosen for *in vitro* translation study. Both dSac12-IA-1 and IA-1-18 cDNA, containing the complete IA-1 open reading frame, were linearized with either *Bam*HI or *Xho*I restriction enzyme. Capped mRNA (Stratagene) was synthesized using either T3 or T7 RNA polymerase to produce either the sense or antisense transcript. One μ g of each transcript was added to a rabbit reticulocyte *in vitro* translation reaction mixture (Promega) in the presence of [35 S]cysteine (Amersham Corp.) at 30 °C for 1 h. Twenty μ l of reaction mixture was run on 7.5% SDS-polyacrylamide gel electrophoresis and fixed with autoradiography enhancer (DuPont NEN). The gel was exposed to the film overnight at room temperature.

Fluorescence In Situ Hybridization.—The procedure utilized has been reported by Tory *et al.* (1992). Briefly, biotinylated IA-1 genomic DNA

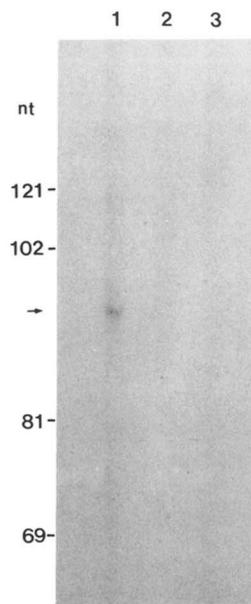


FIG. 5. **RNase protection assay.** A 32 P-labeled antisense riboprobe was prepared by T7 RNA polymerase transcription from a genomic subclone, dSac3BE. The riboprobe was purified from a 6% urea gel, and 1×10^5 cpm riboprobe was mixed with 50 μ g of NCI-H69 RNA (lane 1), HeLa RNA (lane 2), or yeast tRNA (lane 3) for overnight hybridization. The hybridized mixture was subjected to RNase digestion and analysis on a 8% denaturing sequencing gel. An estimated 100-nucleotide (nt) long fragment (lane 1) was protected with small cell lung carcinoma RNA (arrow). Riboprobe size markers were prepared from the same vector.

probe (20 ng/ μ l) was hybridized with metaphase chromosomes overnight at 37 °C in a solution of 50% formamide and 10% dextran sulfate in $2 \times$ SSC. The slides were washed with 50% formamide in $2 \times$ SSC followed by $2 \times$ SSC alone, both at 40 °C. Detection was carried out using avidin-fluorescein isothiocyanate. Chromosome identification was performed following simultaneous staining with Hoechst 33258.

RESULTS AND DISCUSSION

Genomic Structure and Nucleotide Sequence of the IA-1 Gene.—In this study, we report the cloning and sequencing of the IA-1 gene from a human normal liver genomic library. By screening 2×10^5 recombinants with IA-1-34 cDNA probe, we isolated two identical genomic clones (λ 3) containing an approximately 13-kb insert. The λ 3 genomic clone was digested with *Sac*I and subcloned into the pBlueScriptII (sk+) vector. Four subclones, designated pL3S4.9, pL3S0.8, pL3S2.0, and pL3S5.6, were generated (Fig. 1). The IA-1 gene and the 5'-upstream flanking region (2090 bp) were completely sequenced from pL3S0.8, pL3S2.0, and the 5'-region of pL3S5.6. Our experiments showed that the nucleotide sequence derived from normal human genomic DNA was identical to the IA-1 cDNA sequence derived from human insulinoma mRNA (Goto *et al.*, 1992). In addition, we have completed the restriction map of the λ 3 genomic clone by performing restriction enzyme mapping on pL3S5.6 and pL3S4.9. Both sequence analysis and restriction enzyme mapping indicated that the IA-1 gene is not interrupted and, therefore, appears to be an intronless gene (Fig. 1).

The existence of introns has been reported in most of the genes of eukaryotic organisms. However, some important eukaryotic genes, such as seven transmembrane receptors (Allard *et al.*, 1987; Kobilka *et al.*, 1987a, 1987b; Young *et al.*, 1986; Sunahara *et al.*, 1990; Demchyshyn *et al.*, 1992; Nakayama *et al.*, 1985), human interferon- α and - β 1 (Nagata *et al.*, 1980; Lawn *et al.*, 1981), type X collagen (Ninomiya *et al.*, 1986), and histone genes (Stein, 1984) are intronless. The number of genes without introns is small as compared with those with introns.

UPE

-2090 GAGCTCAAAA GAGCAAGGAA ATAACATCAG TTACTTTTAG ATAGAAATAT GTTTATAAAT TTTAGATAAA CTCATTTTAA GATTTCTAAC TTTTGGGGAG
 -1990 ACTAGCACTC ACCACAGCAT TTTTAAAACT AGGGTGAGAC TTTTCAAAAT ATATTTCACT ATATGCATTC TTCCCTTTTC AGTGTGTGAG TTTAAAATATT
 -1890 ACAAACATAA AATTATCAAA TATCTTTTCC TTGATTAACC AACCTGATTA ACCCCTTAAT TAATTAAGCC TAACAACCTGT CCCCATCACA ACTTATGACT
 -1790 CATGTGGACT CTATTTTTTAC ATGTTTAAAT TCTACGCCAG GGTATTTTCA GAGGGAAAAT CCAAAAACAAT TAAATACCTT CAGTGAATTT AAGCGTAAGT
 -1690 AAAACTCAGC AAAATCTGTC TGGTATACAA AGCTTTGCAG AAAAATGTTT CCTTCTCTTT CAATAAAGTG ACATTTTCTT CACAGCATAT GACATTTTAT
 -1590 GTGGCTTGGC TATTTGTTCC CCAGTGTGGC ACGATCAGGG CCACTACGCA CTCGTGGGGT TCTCTGCCCT GTCCCCCGAA GATTTCTGCT TTTGTGTTTC
 -1490 CTCGGGGCTG GCTTGGACAG AGGATGGAGG CAGGCGGCC AGTCTGGGTC CAAACGGAACT GGCAGCGGAG GTGGGGGTGG GGCTGGGGC GGGGGGAATC
 -1390 TTCGCTGTTG TCTCTTCTG GTGGCACAGG GGAGCGGCTT TGTCCTCTCG GAGGCCCTCAG CCTGCCTCAG AGTACAGATT GCCCCCCCTC CCCCCGTTCA
 -1290 GCACGCGTCT CTTTTGCGTC CAGATTGGCC GCGACCGGAG CTCAATAGCA GGAGGTTAAT TTCCTTCACT AAGGTGAAGG GGCACGGCT CCGTGGGGGC
 -1190 TGGCCAGAC AGGCAGCCCC TTTATTTGCG AGCGCCTTGA TTGGAGCCCT TGATTTAGCA TCTGATGTCA ACCGGCAAACT AAAATGCCCG CTCGGAATGA
 -1090 AAATGCATGA GGCCGCGCGG GGAGGAGAAA GCCAACTTCA CTGGGGCGCA CGAGGCCGAC CGCGCGTTTC GGGCTTCGGC CAACTCGACC CGGATTAACC
 -990 GAGCCTTGA CCACGCGGAA GCTCCGGGCG GCCTGGGTCG GCCCGCCAA AGATGCCATT CCCAAGCCTC CAAGAACCCT AAAGTTCAGA AATTCACGTT
 -890 CCTCTGGGAA CCCAGCCCAG CCCCGCCGCG CTCCAATCCC GCCTTTTCGG GAGACTGAGA AGGCCGTGCC AATTCCTTCT CAAACTCGAA AGAAAACCTC
 -790 TCTAGCCCCG TGGCGCGCGG GAGGCTGCGA GCACAAACAT CGCCCTCGGC CACTGCCAGA AGGCCGGGCC CCTGTCCAC ACTTGAACC CCGGGGAACC
 -690 CTTTTGCTTG GCCTCTTGGG TCCAGCGGCC CATCCGTCCA AGGTCCGGGC GGAGCGGTCC GGACCCTGCT GCTCTCTCGG ATTCTGTGTT ATTTCCAAA
 -590 CACCACGCGG ACGCACTGCG CCTCCGCAAC GATCTCCCC GCACCGCCCG CGCCGCCACC CAATCAGCGC GCACAACTTC CCCCTCGGCT
 -490 CCGGCTCGCG GATTGAACCC TCCTGACATA TTTGGGGCCA TTCTTCTCTT TTGTTGCTAT TTTGCTAGCG ACCCGGGGT AATCCCGCG CCGGAGGGGG
 -390 GCGTGCATTG TCGCGTGAT GGACGGGCC ATTTGGCGGC TCCGCGCCCC CCGGAGGAGA GACACAAAGC CCAGGCACGT GCGCCTCCCC ATAGAGAAAGC
 -290 AGCAGACCGT GAAGGGAGGC GGGCGCGGCG GTGTGCCTGG ACCGGGGCGG CGGCGGCGC CGGGCGGGC GACCAGGGGC GCGCGCGGG GCCCCGCGCC
 -190 CTCAGGTACA TCTGCCGAC CTACCGGGCG ACCCCGAGT CCCGCCCCC TTTTGGCGC CCCATCGCC TCCACCCCTG CCAGGCTGAG GAGCTGCGGA
 -90 CCGCTGATT GCCTCCAGG GAAGCGGGAG GCGAGAACA TGGCCCCCT CCCCCGTTAA AAGGGAGCG CTGCCGGGC CCGGGACAGG GACGCGCGTG
 11 CAGGGCGCAG AGCTGGGCG AGCCGTCGCC GCGCCACGC GAGTCCCGCA GCCCGCGCG CCGGGCAATG GGCCGGGGC ACTGAGGGCC GCCGGGGCG
 111 AGCGCGGAG GGGGACCGAG CCAGTGCCGT GCCCTCGGC CCGCGAACA TGCCCGCGG CTTCCTGGT AAGCGCAGCA AG

FIG. 6. Nucleotide sequence of the 5'-end of the human IA-1 gene. Transcription initiation site is assigned to position +1. An asterisk (*) indicates the TATA box-like sequence. Conserved sequences of known regulatory elements, SP-1, T-Ag, AP-1, upstream promoter elements (UPE), and GHF-1 are underlined.

The studies described here show that IA-1, which encodes a zinc finger DNA-binding protein, belongs to this small group of genes that are intronless.

Southern analysis of the IA-1 gene (Fig. 2) revealed a single band (6.5 kb) with *EcoRI* digestion and two bands (3.8 and 0.75 kb) with *HindIII* digestion. These findings are consistent with the restriction enzyme mapping data shown in Fig. 1 and suggest that the IA-1 is probably a single copy gene. DNAs isolated from a total of 16 individuals (not shown) digested with either *EcoRI* or *HindIII* restriction enzymes did not show any evidence of restriction fragment length polymorphisms.

In Vitro Translation Products from IA-1 cDNA and IA-1 Genomic DNA—To show that both IA-1 cDNA and IA-1 genomic DNA can translate the same protein product without RNA splicing, IA-1 cDNA from position +12 to the poly(A) tail (IA-1-18) (Goto *et al.*, 1992) and an IA-1 genomic DNA clone (dSac12-IA-1) containing 304 bp 5'-upstream from the ATG start codon were used as DNA templates for synthesizing RNA transcripts. Using the rabbit reticulocyte system for *in vitro* translation, a prominent protein product with an estimated M_r of 61,000 was produced by both IA-1 cDNA and the intronless IA-1 genomic DNA (Fig. 3). Antisense transcripts revealed no protein product. This demonstrates that the intronless IA-1 gene is capable of translating into the same protein product as the IA-1 cDNA.

Mapping of the IA-1 Gene to Chromosome 20—The fact that IA-1 is a sensitive marker for small cell lung cancer (Lan *et al.*, 1993) and has a restricted tissue distribution (Goto *et al.*, 1992) makes it important to determine its chromosomal localization. Using fluorescence *in situ* hybridization, a total of 46 metaphase cells were examined. In 42 of these cells, from one to

four chromatids were labeled at 20p11.2 (Fig. 4). No significant background was observed at other chromosomal positions. These results suggest that the IA-1 gene is a single copy gene located on the short arm of chromosome 20. Several hereditary diseases have been mapped to the short arm of chromosome 20. Thus far, only one disease, arteriohepatic dysplasia, has been localized specifically to the p11.2 region (Tebebi *et al.*, 1992; Simpson, 1988).

Characterization of 5'-Upstream Sequence of the IA-1 Gene—The transcription initiation site of the IA-1 gene was determined by RNase protection assay using total RNA isolated from both a small cell lung carcinoma cell line (NCI-H69) and HeLa cells. A gel-purified antisense riboprobe extending from 459 to 60 bp upstream of the ATG start site was hybridized with 50 μ g of NCI-H69 RNA, HeLa cell RNA, or yeast tRNA. As shown in Fig. 5, an estimated 100-nucleotide long fragment (lane 1) was protected specifically with small cell lung carcinoma RNA, which strongly expressed IA-1 message, whereas HeLa RNA and tRNA did not protect the probe (lanes 2 and 3). The transcription start site was estimated at 159 bp upstream of the ATG start codon (designated as +1 in Fig. 6), which is 12 bp longer than the IA-1 cDNA reported previously (Goto *et al.*, 1992). At the vicinity of the translation start site, there is a stretch of GC-rich sequence (>80%), which might form secondary structure that could hamper the proper primer extension analysis. In fact, we made several primers but were unable to extend them to the correct transcription initiation site as determined by RNase protection assay.

The 5'-upstream DNA sequence, extending 2090 bp upstream from the transcription start site, was determined by double-stranded sequencing of pL3S2.0 and pL3S0.8 genomic

subclones (Fig. 6) and compared with known sequences in the GenBank DNA data base (Pearson and Lipman, 1988). No significant similarity was found except for a number of conserved regulatory elements. The classic TATA box (McKnight and Kingsbury, 1982) typically located between positions -25 and -35 bp was not found; however, an AT-rich sequence located between positions -29 and -34 bp, which might serve as a TATA box, was observed. The so-called CAATT box, which is often found at -80 bp, was absent. Three putative transcription factor SP-1 binding sites (GC boxes) located between positions -220 and -250 bp (Kadonaga *et al.*, 1987) and three additional GC boxes further upstream in reverse orientation also were found. In this context, the promoters of several housekeeping genes that lack the characteristic TATA and CAATT boxes (Melton *et al.*, 1984; Reynolds *et al.*, 1984; Valerio *et al.*, 1985) are known to have a similarly high GC content. Further upstream, three SV40 T antigen binding sites (Jones and Tjian, 1984) and an AP-1 (Lee *et al.*, 1987a, 1987b) binding site were found. The SV40 T antigen binding site has been shown to be both a positive regulator of viral replication and a negative regulator of early gene transcription (Fanning and Knippers, 1992). The AP-1 binding site binds a transcription factor, AP-1, and its activities may be modulated by treatment of cells with the tumor promoter, 12-O-tetradecanoylphorbol-13-acetate (Angel *et al.*, 1987; Lee *et al.*, 1987a).

Two β -cell-specific upstream promoter elements, characteristic of the upstream glucokinase promoter, are located at positions -2075 and -1807 bp of IA-1. Gel mobility shift experiments have shown that a β -cell factor isolated from nuclear extracts of murine insulinoma cell lines (HIT M2.2.2 and β -TC-3), but not from the pituitary tumor cell line (AtT-20) or the fibroblast cell line (NIH-3T3), binds to the upstream promoter elements (Shelton *et al.*, 1992). The β -cell factor is the same factor that binds to similar elements, termed CT boxes, in the insulin promoter. Additionally, at position -1931 bp, there is a conserved prolactin/growth hormone gene recognition element GHF-1 (TATATGCAT), which is the binding site of the pituitary-specific positive transcription factor, Pit-1 (Ingraham *et al.*, 1988). It has been suggested that the binding of the Pit-1 factor to this sequence may confer a characteristic cellular phenotype in the anterior pituitary.

The IA-1 gene is expressed in tumors of neuroendocrine origin, including insulinomas and pituitary tumors. The presence of several different tissue-specific regulatory elements in the upstream region of the IA-1 gene may result in the expression of IA-1 in different neuroendocrine cells responding to different tissue-specific transcription factors. Currently, we are attempting to characterize the promoter and other regulatory elements of the IA-1 gene as well as tissue-specific transcription factors present in various IA-1-expressing cells, including small cell

lung carcinoma, pheochromocytoma, and medullary thyroid carcinoma.

Acknowledgments—We thank Dr. Y. Goto for screening the genomic library and Drs. P. Zhou, M. F. Young, and R. R. Franks for advice and critical reading of the manuscript. We gratefully acknowledge the excellent editorial help of Eloise Mange.

REFERENCES

- Allard, W. J., Sigal, I. S., and Dixon, A. F. (1987) *Nucleic Acids Res.* **15**, 10604
- Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P., and Karin, M. (1987) *Cell* **49**, 729–739
- Demchysyshyn, L., Sunahara, R. K., Miller, K., Teitler, M., Hoffman, B. J., Kennedy, J. L., Seeman, P., VanTol, H. H. M., and Niznik, H. B. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5522–5526
- Fanning, E., and Knippers, R. (1992) *Annu. Rev. Biochem.* **61**, 55–85
- Goto, Y., DeSilva, M. G., Toscani, A., Prabhakar, B. S., Notkins, A. L., and Lan, M. S. (1992) *J. Biol. Chem.* **267**, 15252–15257
- Ingraham, H. A., Chen, R., Mangalam, H. J., Elsholtz, H. P., Flynn, S. E., Lin, C. R., Simmons, D. M., Swanson, L., and Rosenfeld, M. G. (1988) *Cell* **55**, 519–529
- Jones, K. A., and Tjian, R. (1984) *Cell* **36**, 155–162
- Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51**, 1079–1090
- Kobilka, B. K., Matsui, H., Kobilka, T. S., Yang-Feng, T. L., Francke, U., Caron, M. G., Lefkowitz, R. J., and Regan, J. W. (1987a) *Science* **238**, 650–656
- Kobilka, B. K., Frielle, T., Dohlman, H. G., Bolanowski, M. A., Dixon, R. A. F., Keller, P., Caron, M. G., and Lefkowitz, R. J. (1987b) *J. Biol. Chem.* **262**, 7321–7327
- Lan, M. S., Russell, E. K., Lu, J., Johnson, B. E., and Notkins, A. L. (1993) *Cancer Res.* **53**, 4169–4171
- Lawn, R. M., Adelman, J., Franke, A. E., Houck, C. M., Gross, M., Najarian, R., and Goeddel, D. V. (1981) *Nucleic Acids Res.* **9**, 1045–1052
- Lee, W., Mitchell, P., and Tjian, R. (1987a) *Cell* **49**, 741–752
- Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987b) *Nature* **325**, 368–372
- McKnight, S. L., and Kingsbury, R. (1982) *Science* **217**, 316–324
- Melton, D. W., Konecki, D. S., Brennand, J., and Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2147–2151
- Nagata, S., Mantei, N., and Weissmann, C. (1980) *Nature* **287**, 401–408
- Nakayama, N., Miyajima, A., and Arai, K. (1985) *EMBO J.* **4**, 2643–2648
- Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenymer, T., and Olsen, B. R. (1986) *J. Biol. Chem.* **261**, 5041–5050
- Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2444–2448
- Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L., and Luskey, K. L. (1984) *Cell* **38**, 275–285
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, pp. 9.16–9.19, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Shelton, K. D., Franklin, A. J., Khor, A., Beechem, J., and Magnuson, M. A. (1992) *Mol. Cell. Biol.* **12**, 4578–4589
- Simpson, N. E. (1988) *J. Med. Genet.* **25**, 794–804
- Stein, G. S., Stein, J. L., and Marzluff, W. F. (1984) *Histone Genes: Structure, Organization, and Regulation*, pp. 397–455, John Wiley & Sons, Inc., New York
- Sunahara, R. K., Niznik, H., Weiner, D., Stormann, T., Brann, M., Kennedy, J., Glernter, J., Rozmahel, R., Yang, Y., Israel, Y., Seeman, P., and O'Dowd, B. (1990) *Nature* **347**, 80–83
- Teebi, A. S., Murthy, D. S., Ismail, E. A., and Redha, A. A. (1992) *Am. J. Med. Genet.* **42**, 35–38
- Tory, K., Latif, F., Modi, W., Schmidt, L., Hui-Wei, M., Li, H., Cobler, P., Orcutt, M. L., Delisio, J., Geil, L., Zbar, B., and Lerman, M. I. (1992) *Genomics* **13**, 275–286
- Valerio, D., Duyvesteyn, M. G. C., Dekker, B. M. M., Weeda, G., Berkvens, T. M., van der Voorn, L., van Ormondt, H., and van der Eb, A. J. (1985) *EMBO J.* **4**, 437–443
- Whittaker, P. A., and Southern, E. M. (1986) *Gene (Amst.)* **41**, 129–134
- Young, D., Waitches, G., Birchmeier, C., Fasano, O., and Wigler, M. (1986) *Cell* **45**, 711–719