Characterization of the Promoter Region of the Human Insulin Receptor Gene

EVIDENCE FOR PROMOTER ACTIVITY*

Eiichi Araki†, Fumio Shimada‡, Haruo Uzawa§, Masataka Mori¶, and Yousuke Ebina†‡

From the †Institute for Medical Genetics and ‡Department of Metabolic Medicine, Kumamoto University Medical School, Kumamoto 862, Japan.

Recombinant clones containing the promoter region of the human insulin receptor gene were isolated from genomic libraries derived from nondiabetic persons. A 1.5-kilobase pair fragment of the 5'-flanking region was sequenced. One transcriptional start site, located at 203 bases upstream from the start of translation was identified by nuclease S1 mapping and the primer extension experiment using the human insulin receptor mRNA. The bacterial chloramphenicol acetyltransferase assay revealed that a 573-base pair fragment immediately preceding the ATG has promoter activity and that the transcript initiates from the normal start site of the insulin receptor gene in the COS cells.

The promoter region contains neither a "TATA box" nor a "CAAT box," has an extremely high G + C content, and contains seven central components of potential Sp1 binding sites (GGCGGG or CCGCCC). These features are common to those found in the regulatory regions of a class of constitutively expressed "housekeeping" genes. A comparison between the promoter sequence of the human insulin receptor and those of other "housekeeping" genes revealed the presence of homologous sequences among these genes, in addition to the potential Sp1 binding sites.

Insulin initiates diverse biological responses by binding to its receptor, an integral membrane glycoprotein composed of two α (Mr = 135,000) and two β (Mr = 95,000) subunits linked by disulfide bonds (1-4). Insulin binding to the β subunit of the receptor results in stimulation of intrinsic tyrosine kinase activity of the β subunit (5, 6), which is essential for key physiological responses of this peptide hormone (7-9). The receptor-insulin complexes are subsequently internalized into vesicles, and most of the internalized receptors are recycled back to the plasma membrane, while some are destroyed by association with lysosomes (reviewed in Ref. 1). Chronic exposure of cells to insulin leads to a net loss of receptors from the plasma membrane (down-regulation) (10-12).

Although the insulin receptor (IR) is considered to be one of the "housekeeping" proteins essential for cell growth and is usually synthesized at a low level in all cells (13), the number of cell surface receptors greatly increases under certain conditions. After treatment with dexamethasone and isobutylmethylxanthine, mouse 3T3-L1 fibroblasts differentiate into adipocytes in a process which leads to about 10-fold induction of the synthesis of IR mRNA (4). Expression of the Drosophila IR gene is developmentally regulated; the expression of the gene is greatest between 8 and 12 h of embryogenesis (14, 15). The IR homologous down-regulation was noted to be associated with increased IR biosynthesis (16), and glucocorticoids induce the synthesis of IR (17).

The regulation of IR gene expression also seems to be important from the clinical aspect. Diabetes mellitus is caused either by a deficiency in insulin or by insensitivity of the target cells to insulin. The latter defect causes the majority of diabetes cases, and many of them are hereditary (18). It seems plausible that an abnormal structure or expression of the IR gene which causes a decreased functional IR of cells contributes to the insensitivity to insulin, and diabetes mellitus ensues. A decrease in the number of IR was noted in a diabetic patient with severe insulin resistance (19).

To elucidate the basis of regulation of IR gene expression and to provide a foundation upon which to examine the relationship between diabetes and the promoter abnormalities, we identified and characterized the promoter region of the IR gene from nondiabetic persons.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of Human Genomic Clones—Genomic DNAs were obtained from leukocytes of two nondiabetic persons (20). The genomic libraries were constructed in the λ phage vector EMBL4 from a partial Sau3A1 digest of DNA, according to the protocol of the suppliers (Promega Biotec, Madison, WI). Approximately 10^6 phages from each library were screened by plaque hybridization using an EcoRI-EcoRV cDNA fragment (200 bp) containing the 5' sequence of the human IR cDNA (3). Phage DNAs of positive clones were subcloned into pUC18 vectors or M13 phages, and the nucleotide sequences of both strands were determined using the dideoxynucleotide chain termination procedure (21). The sequence of the G-C-rich region was confirmed by the method of Maxam and Gilbert (22).

Analysis of the 5' End of the Human IR mRNA—Poly(A)^+ RNA was prepared from human IM9 cells by guanidine thiocyanate extraction (23) followed by oligo(dT)-cellulose chromatography. For 5' nucleotide mapping, a 1450-bp DNA fragment excised from pE44 with NcoI and SphI (see Fig. 1) was used. The probe was end-labeled with [γ-32P]ATP and polynucleotide kinase and digested with HindIII which yielded a 573-nucleotide HindIII-NcoI DNA with a labeled 5' terminal nucleotide of the NcoI site. The probe was added to 4 μg of

*This work was supported by Grantes-in-Aid from the Ministry of Education, Science and Culture of Japan and from the Mochida Memorial Foundation for Medical and Pharmaceutical Research (to Y. E.). 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.

†To whom correspondence should be addressed: Institute for Medical Genetics, Kumamoto University Medical School, Kuhonji 4-24-1, Kumamoto 862, Japan.

‡The abbreviations used are: IR, insulin receptor; bp, base pair(s).
poly(A)* RNA in 80% formamide (Bethesda Research Laboratories) supplemented with $1 \times$ Pipes buffer ($24, 25$). The mixture was heated at $80\, ^\circ C$ for 10 min. Hybridization was done at $60\, ^\circ C$ for 12 h. S1 nuclease digestion was performed at $30\, ^\circ C$ with 500 units of the nuclease for 90 min. S1 digestion products were run on a sequencing gel alongside a dideoxy sequencing ladder priming from the same NcoI site that bordered the S1 probe.

For primer extension analysis, a 23-mer single-stranded DNA, 5'-GCCCGGGGGTTACATGCTCCGAG-3', corresponding to nucleotide 135-157 (Fig. 2) was synthesized, using a DNA synthesizer, and end-labeled with $[\gamma-^3P]ATP$ and T4 polynucleotide kinase. 4 pg of poly(A)* mRNA and 5 pmol of the end-labeled primer were co- precipitated with ethanol. The pellet was dissolved in 12.5 pmol of 500 mM EDTA, denatured at 90 °C for 3 min, and 2.5 pmol of 500 mM Tris-HCl (pH 8.3), 500 mM KCl, 50 mM MgCl, and 400 pmol of dithiothreitol was added. Hybridization was carried out at $37\, ^\circ C$ for 10 min. The reaction mixture for primer extension contained 6 pmol of poly(A)* RNA/primer annealing mixture mentioned above, 3 ml of 2.5 ml each of four kinds of deoxynucleotide triphosphates, and 14 units of reverse transcriptase. The mixture was incubated at $44\, ^\circ C$ for 30 min. After addition of 10 pg of tRNA as a carrier, the reaction mixture was extracted with phenol and precipitated with ethanol. The primer- extended DNA products were run on a sequencing gel alongside a dideoxy sequencing ladder priming for the same primer for the primer extension experiments.

**Plasmid Constructions for Chloramphenicol Acetyltransferase Assay.—** A 3.5-kilobase SpHl-Xbal fragment of pEA22 was subcloned into the HindIII site of pUC18 to construct pEA4 (see Fig. 4). After NcoI digestion of pEA4, the DNA fragments were treated with mung bean nuclease to make a blunt end, and the HindIII linker was ligated. The mixture was partially digested with HindIII. The proximal 575-bp HindIII-HindII fragment and the entire 1462-bp HindIII-HindIII fragment immediately preceding the ATG codon were purified. These fragments were ligated into an unique restriction site that bordered the S1 probe.

**RESULTS**

**Isolation of 5'-Specific Human IR Genomic Clones—** Four independent genomic clones containing the 5' end of the IR gene were isolated from two different genomic DNA libraries by screening with the 200-bp probe containing the 5' sequence of the human IR cDNA (3) (Fig. 1). Since the restriction maps of these clones are identical, we chose one clone, pEA22, to determine the sequence of the IR promoter region. Sequence analysis of the 1450-bp SpHl-NcoI fragment and the adjacent 3' region from pEA22 revealed that the first exon of IR gene is located in this area; this first exon contains the 203-bp 5'-untranslated region, 51 bp coding for the signal sequence and 48 bp coding for the NH2-terminal portion of the mature IR (Fig. 2). Since a discrepancy in the nucleotide sequence of the 5'-untranslated region was found between the IR cDNA sequence reported in the literature (3) and the IR genomic DNA sequence shown in Fig. 2, the IR clone, IR8 (3), was sequenced again by the method of Maxam-Gilbert and using the dideoxy sequencing method. Errors in the sequence of the 5'-untranslated region of the IR cDNA thus became apparent. The redetermined sequence of the IR cDNA was confirmed to be identical with that of the IR genomic DNA shown in Fig. 2.

**Determination of Transcription Initiation Site by Primer Extension and Nuclease S1 Mapping**—A combination of S1 nuclease protection assay and primer extension reaction was used to identify the probable transcription initiation site on the human IR gene. A 573-bp HindIII-NcoI DNA fragment extending from the nucleotide +206 to the nucleotide +367 was used for the S1 mapping experiments (see Fig. 2). S1 nuclease treatment of the hybrids between the DNA fragment and mRNA from human IM9 cells produced protected fragments of around 206 nucleotides long corresponding to the initiation site at nucleotide 1 (Fig. 3A). No other protected fragments and protected full-length probe were seen in this autoradiogram (data not shown). Since a certain degree of microheterogeneity, which was reported to be inherent to the S1 nuclease assay (28, 29), was seen in the protected DNA bands, we used another approach, primer extension reaction. A single-stranded 23-mer DNA corresponding to the nucleotide sequence starting from +135 to +157 nucleotide (see Fig. 2) was synthesized and used as a primer in the DNA elongation reaction as described under "Experimental Procedures." Total poly(A)* RNA isolated from IM9 cells was used as the template. The extended DNA product was 157 nucleotides in length (Fig. 3B); the result agrees with that of S1 mapping analysis. Another 417-nucleotide extended product was seen in the autoradiogram of the primer extension experiment, but the corresponding band was not detected in the autoradiogram of the primer extension experiment, but the corresponding band was not detected in the S1 mapping experiment (data not shown). These results indicate that the transcriptional initiation site of the IR gene localizes at thymine 1 of the sequence shown in Fig. 2. Total poly(A)* RNA isolated from the human placenta was also used as the template for the primer extension experiment. The extended DNA products were similar to those shown in Fig. 3B (data not shown).

Typically a region between 20 and 30 bases upstream is required for normal transcription initiation at the cap site. In most eukaryotic genes heretofore studied, this region contains a "TATA box" (30). More distal elements in the promoter region, typically extending to 100 or more bases from the cap site, are important to control the frequency of initiation, and a second consensus sequence, "CAAT box," is often located in this distal promoter region (31). However, the promoter
region for the human IR gene lacks a “TATA box” or a “CAAT box.” This region has a high G + C content (77%; +1 to −371 region) and seven potential sites for the cellular transcription factor Sp1 (hexanucleotide: GGGCGG or CCGGCC) at positions −199, −226, −230, −391, −401, −406, and −411. Two dyad symmetries exist at positions −37 to −57 and −53 to −83. In these dyad symmetric regions, we found sequences resembling those in the promoter regions of several housekeeping genes (see “Discussion”). These are two sets of direct repeats, CCCCGG(C)GCGCA, and GCCCGCCCGCCA, in the promoter region (Fig. 2).

Evidence for Promoter Activity—To determine whether the 1450-bp SpI-NcoI fragment does indeed contain a functional active promoter, several constructs were made with DNA fragments obtained by addition of HindIII linkers at SpI and NcoI sites. These fragments were inserted at the unique HindIII site in the expression vector pSVOCAT (26) (Fig. 4). Since there is a putative ribosome binding site around the ATG codon (nucleotide 199–208) of the human IR gene, four nucleotides in this region were deleted by digestion with mung bean nuclease to prevent the start of translation from this ATG codon. As shown in Fig. 2, there are two clusters of potential Sp1 binding sites located from the nucleotide −196 to −232 and −389 to −414. The orientations of the three Sp1 binding sites on the 573-bp HindIII-NcoI fragment are opposite those of the other four Sp1 binding sites on the 828-bp SpI-HindIII fragment. A number of promoters which contain actual or potential Sp1 binding sites have been described (32). Among them, the in vitro transcription of SV40 early, human metallothionein Iα, and mouse hypoxanthine phosphoribosyltransferase (M-HPR) promoters are underlined. The locations of these elements in each promoter region are: human 3-phosphoglycerate kinase, −62 to −72 (35); human epidermal growth factor receptor, +29 to +40 (36); human adenosine deaminase, −28 to −41 (39); mouse hypoxanthine phosphoribosyltransferase −32 to −44 (40) from the first transcription initiation site. The location of the opposite strand of the 23-mer primer for the primer extension experiment (Fig. 3B) is indicated by a wavy line.
Human Insulin Receptor Gene Promoter

A

Human  Insulin Receptor Gene Promoter

16189

A

I.*

I

Hind 111

B

GATC

Ncol

0A

IC A

"A

I

I

Probe 573nt.

- mRNA

Protected Probe

206nt.

mRNA

Primer

Extension

- mRNA

Primer

Extended DNA

I57 nt.

FIG. 3. Determination of the 5′ end of the human IR mRNA by S1 nuclease mapping (A) and primer extension (B). Poly(A)⁺ RNA from human IM9 cells was analyzed. A, the probe was a HindIII-NcoI fragment (+202 to −371) ³²P-labeled at the 5′ end of the NcoI site. Hybridization and S1 nuclease digestion were carried out as described under “Experimental Procedures.” B, the primer was a synthetic oligonucleotide (23-mer, complementary to position +135 to +157) ³²P-labeled at the 5′ end. The primer extension experiment was carried out as described under “Experimental Procedures.” To compare the products directly with the genomic nucleotide sequence, the arrows indicate the position assigned to the 5′ end of the mRNA.

(25). Extracts of COS cells transfected with the pEA14 or pEA16 DNAs promoted the formation of acetyl chloramphenicol, whereas those from cells transfected with pSV0CAT (without promoter) had little activity (Fig. 5A). pSV2CAT, a positive control DNA which contains an SV40 early promoter and enhancer (26), produced a relatively high level of chloramphenicol acetyltransferase activity. A low chloramphenicol acetyltransferase activity was detected in the extracts of COS cells transfected with pEA15 which has the proximal 575-bp HindIII-HindIII fragment of the IR promoter in the opposite orientation to chloramphenicol acetyltransferase gene.

DISCUSSION

We have localized the IR gene “promoter” in cloned human genomic DNA. The IR gene does not possess typical TATA or CAAT boxes. The 5′-flanking region of the IR gene, therefore, differs from the corresponding region of most other eukaryotic genes but does resemble the promoters of the housekeeping genes such as those for human 3-phosphoglycerate kinase (35), human epidermal growth factor receptor (36), human hypoxanthine phosphoribosyltransferase (37), hamster 3-hydroxy-3-methylglutaryl coenzyme A reductase (38), human adenosine deaminase (39), and mouse hypoxanthine phosphoribosyltransferase (40). Common features of
Chloramphenicol acetyltransferase activity in COS cell lysates (A) and S1 analysis of mRNA from COS cells transiently transfected with pEA14 (B).

A.

1. COS cells transfected with the chloramphenicol acetyltransferase (CAT) vectors described in Fig. 4 were lysed and assayed for chloramphenicol acetyltransferase enzymatic activity as described under "Experimental Procedures." The two adjacent lanes represent independent transfections of each plasmid DNA. CM, chloramphenicol; B, poly(A)+ RNA from COS cells transfected with pEA14 was analyzed. The probe was a HindIII-HindII1 fragment (575 bp) 32P-labeled at the 5' end of HindIII site adjacent to the chloramphenicol acetyltransferase gene. The hybridization and the S1 nuclease digestion were carried out as described under "Experimental Procedures." To compare the products directly with the genomic nucleotide sequence, the 5' end of the probe was arranged to coincidence with the 5' end of the primer of the dideoxynucleotide sequencing. The thick arrow indicates the position assigned to the 5' end of the mRNA.

B.

Fig. 5. Chloramphenicol acetyltransferase activity in COS cell lysates (A) and S1 analysis of mRNA from COS cells transiently transfected with pEA14 (B). A, COS cells transfected with the chloramphenicol acetyltransferase (CAT) vectors described in Fig. 4 were lysed and assayed for chloramphenicol acetyltransferase enzymatic activity as described under "Experimental Procedures." The two adjacent lanes represent independent transfections of each plasmid DNA. CM, chloramphenicol. B, poly(A)+ RNA from COS cells transfected with pEA14 was analyzed. The probe was a HindIII-HindIII fragment (575 bp) 32P-labeled at the 5' end of HindIII site adjacent to the chloramphenicol acetyltransferase gene. The hybridization and the S1 nuclease digestion were carried out as described under "Experimental Procedures." To compare the products directly with the genomic nucleotide sequence, the 5' end of the probe was arranged to coincidence with the 5' end of the primer of the dideoxynucleotide sequencing. The thick arrow indicates the position assigned to the 5' end of the mRNA.

the promoter regions of these genes are lack of typical TATA and CAAT boxes, high G/C content upstream of the cap site, and existence of potential Sp1 binding sites (G-C boxes; GGGCGG or its inverted form CCGCCC) with varying distances from the cap sites (35–40). Like the corresponding regions of other housekeeping genes, the 5' flanking region of the IR gene is G + C-rich. Within this area of the IR gene, there are seven repeats of the potential Sp1 binding site. Transcription factor Sp1 was recently purified (33) and was shown to enhance transcription by RNA polymerase II 10–50-fold from several viral and cellular promoters that contain at least one properly positioned Sp1 binding site (33, 34). Transcription of the human IR gene may be responsive to the Sp1 factor. Although the three potential Sp1 binding sites on the proximal 575-bp HindIII-HindIII fragment seem to be sufficient for the expression of the chloramphenicol acetyltransferase gene in transiently transfected COS cells (Fig. 5A), we did not exclude the possibility that the other four Sp1 binding sites are necessary under certain conditions.

Multiple initiation sites for the transcription have been noted in several housekeeping genes lacking the TATA box (35–40), which may determine the precise point of initiation, but the initiation of IR mRNA transcription seems to occur at one site (Fig. 3). Induction of the IR by hydrocortisone has been reported (17). In the 1555-bp region of the IR promoter, we found no good consensus sequence of the glucocorticoid receptor binding site (C(C/G)TGNNTCACTNTGTCCT) (41) except for two elements which are weakly homologous: GCCGCGGTTCCTTGTC, −155 to −171 and TCCCTCCATGAGTTCT, −1137 to −1153 (Fig. 2).

In the promoter regions of those housekeeping genes mentioned above, no consensus sequence except for the potential Sp1 binding sequence has been found. However, we detected some homologous sequences in the human IR promoter and promoters of the human 3-phosphoglycerate kinase (35), the human epidermal growth factor receptor (36), the human adenosine deaminase (39), and mouse hypoxanthine phosphoribosyltransferase (40) genes. These homologous sequences are localized in regions of the two inverted repeats of the IR promoter located at −37 to −57 and −53 to −83 upstream of the cap site (Fig. 2). Inverted repeats present just upstream of the cap site in the IR promoter suggest the binding of transcriptional factor(s) to the regions. Interestingly, the homologous elements of the human 3-phosphoglycerate kinase and the human adenosine deaminase promoters shown in Fig. 2 exist in their own inverted repeats of those promoters (3-phosphoglycerate kinase, −66 to −100 (35); adenosine deaminase, −28 to −57 (39) from the first cap site). The homology in the position of the inverted repeats in these promoters may indicate the relevance of their transcriptional regulations.

Some of the noninsulin-dependent diabetes mellitus may be caused by impaired expression of the IR gene on probably both allelic and subsequent decrement of IR on the cell surface, which, in turn, makes the cells insensitive to insulin. Our data make feasible a comparison of the upstream sequence of the IR gene in diabetics with the sequence of a nondiabetic shown in Fig. 2.

Acknowledgments—We thank Drs. S. Yoshida and H. Makino (Chiba University) and C. Setoyama and M. Takiguchi (Kumamoto University) for helpful discussions, and Drs. N. Takahashi and T. Maekawa (Tokyo University) and Y. Fukumaki (Kyushu University) for the plasmids, Dr. Y. Haraguchi (Kumamoto University) for the library, and M. Ohara (Kyushu University) for comments on the manuscripts. We are especially grateful to Dr. W. J. Rutter (University of California) for generously allowing us use of the IR cDNA for the screening of the gene.

REFERENCES

Human Insulin Receptor Gene Promoter