Human Skeletal Muscle Insulin Receptor Substrate-1

Characterization of the cDNA, Gene, and Chromosomal Localization

EIICHI ARAKI, XIAO-JIAN SUN, BURRITT L. HAAG, III, LEE-MING CHUANG, YITAO ZHANG, TERESA L. YANG-FENG, MORRIS F. WHITE, AND C. RONALD KAHN

Insulin receptor substrate-1 is a major substrate of insulin receptor Tyr kinase. We have now cloned the IRS-1 cDNA from human skeletal muscle, one of the most important target tissues of insulin action, localized and cloned the human IRS-1 gene, and studied the expression of the protein in Chinese hamster ovary cells. Human IRS-1 cDNA encodes a 1242 amino acid sequence that is 88% identical with rat liver IRS-1. The 14 potential Tyr phosphorylation sites include 6 Tyr-Met-X-Met motifs and 3 Tyr-X-X-Met motifs that are completely conserved in human IRS-1. Human IRS-1 has >50 possible Ser/Thr phosphorylation sites and one potential ATP-binding site close to the NH_2-terminal. The human IRS-1 gene contains the entire 5′-untranslated region and protein coding region in a single exon and was localized on chromosome 2 q36–37 by in situ hybridization. By Northern blot analysis, IRS-1 mRNA is rare and consists of two species of 6.9 and 6 kilobase. By using quantitative polymerase chain reaction after reverse transcription of total RNA from human fetal tissues, IRS-1 mRNA could be identified in all tissues. When human IRS-1 cDNA was expressed in Chinese hamster ovary cells, the protein migrated between 170,000–180,000 M_r in sodium dodecyl sulfate-polyacrylamide gel electrophoresis and was rapidly Tyr phosphorylated upon insulin stimulation. Thus, IRS-1 is widely expressed and highly conserved across species and tissues. Compared with rat protein, human IRS-1 contains more potential Ser/Thr phosphorylation sites and only one nucleotide binding site. The entire protein coding sequence is contained within a single exon. *Diabetes* 42:1041–54, 1993

From the Research Division at Joslin Diabetes Center, the Department of Medicine at Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts; and the Department of Genetics, Yale University School of Medicine, New Haven, Connecticut.

Address correspondence and reprint requests to Dr. C. Ronald Kahn, Joslin Diabetes Center, One Joslin Place, Boston, MA 02215.

Received for publication 11 September 1992 and accepted in revised form 11 February 1993.

IRS-1, insulin receptor substrate-1; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; YMXM, Tyr-Met-X-Met; YXXM, Tyr-X-X-Met; SH2 and SH3, *src* homology 2 and 3; PtdIns-3, phosphatidylinositol-3; SDS, sodium dodecyl sulfate; NIDDM, non-insulin-dependent diabetes mellitus; CHO, Chinese hamster ovary; kb, kilobase; SSC, sodium chloride-sodium citrate; PCR, polymerase chain reaction; hIR, human insulin receptor; FBS, fetal bovine serum; PMSF, phenylmethylsulfonyl fluoride; DTT, dithiothreitol; PVDF, polyvinylidene difluoride; cpm, cycles per minute; TBE, 1 M tris base with 1 M boric acid and 20 mM EDTA; bp, base pair; GXGXXG, Gly-X-Gly-X-X-Gly; CHO/hIR, Chinese hamster ovary cells overexpressing the human insulin receptor; α-CT-IRS-1, anti–COOH-terminal insulin receptor substrate-1; α-bIRS-1, anti-baculovirus insulin receptor substrate-1.

Insulin initiates diverse biological effects by binding to the α-subunit of the insulin receptor, which, in turn, results in rapid autophosphorylation of the receptor β-subunit and activation of the Tyr kinase activity of the receptor (1–4). In most, if not all, cells and tissues, activation of the insulin receptor induces Tyr phosphorylation of a cytoplasmic protein that migrates as a broad band between 165,000–185,000 M_r on SDS-PAGE, originally termed pp185 (5–7). Activation of the insulin receptor Tyr kinase and stimulation of Tyr phosphorylation of pp185 are decreased in cells expressing mutated or kinase-deficient insulin receptors, and this correlates with a loss in insulin action (8–11). Furthermore, mutant insulin receptors in which Tyr 960 has been replaced by Phe undergo near-normal Tyr phosphorylation, but fail to phosphorylate pp185 and mediate insulin bioeffects (12). Thus, pp185 appears to be a primary substrate of insulin receptor Tyr kinase in vivo and its phosphorylation is linked to insulin action.

Previously, we purified pp185 from rat liver and, based on partial sequence, cloned the major component of this protein from a rat liver cDNA library (7,13). We termed this protein IRS-1. Rat liver IRS-1 represents a new class of signaling molecules with 2 potential nucleotide binding sites, 14 putative Tyr phosphorylation sites, and 47 possible sites of Ser/Thr phosphorylation. Among the potential Tyr phosphorylation sites, 6 are in a Tyr-Met-X-Met (YMXM) motif and 3 are in a Tyr-X-X-Met (YXXM) motif. These two motifs occur in some other Tyr kinases

and have been reported to bind to the SH2 domain of the 85,000-*M*, subunit of the PtdIns-3 kinase, possibly mediating signal transduction in several pathways (14). Recently, Nishiyama and Wands (15) reported cloning of a gene for a similar protein from a human hepatocellular carcinoma. Thus far little is known of possible heterogeneity of IRS-1 among tissues; however, anti-phosphotyrosine blots of SDS gels, performed on extracts of different tissues, reveal a broad band and variable migration, which suggests either protein heterogeneity or differences in the phosphorylation state in different tissues (5,6,16–18). Because skeletal muscle is one of the most important peripheral target tissues for insulin action and is a primary site of insulin resistance in NIDDM, we have cloned the human skeletal muscle IRS-1 cDNA, characterized its gene structure and chromosome localization, and determined its expression in various human tissues.

RESEARCH DESIGN AND METHODS

Two cDNA libraries (oligo dT primed) were made in the λ-*Zap*II plasmid by Stratagene (La Jolla, CA) using poly(A)$^+$ RNA from skeletal muscle samples taken from a nondiabetic adult (M8) and from an adult with NIDDM (M12). A λ-gt 11 human fetal skeletal muscle cDNA library (oligo dT primed) was a gift from L. Kunkel (Harvard Medical School, Boston, MA). The human genomic library EMBL3A was from S.H. Orkin (Children's Hospital, Boston, MA). Bluescript vectors were from Stratagene, and the pCMV-His vector, which contains the histodinol resistance gene, was a gift from M. Birnbaum (Harvard Medical School). Restriction enzymes were from New England Biolabs (Beverly, MA). Reagents for SDS-PAGE were from Bio-Rad (Richmond, CA). Human fetal tissue RNA was a gift from Dr. C. Morton (Brigham and Women's Hospital, Boston, MA). CHO cells and CHO cells overexpressing the hIR (1×10^6 receptors/cell) were as described previously (19). Antibodies were prepared to the COOH-terminal sequence (Thr-Tyr-Ala-Ser-Ile-Asn-Phe-Gln-Lys-Gln-Pro-Glu-Asp-Arg-Gln) of rat IRS-1 (α-CT-IRS-1) or to the intact protein produced in a baculovirus expression system (α-bIRS-1) by injection of New Zealand white rabbits and purified on protein A-sepharose columns (Pharmacia, Piscataway, NJ). Antibodies to phosphotyrosine were prepared and purified as described previously (6).

Cloning and sequencing of human IRS-1 cDNA. Plaques (~10^6) from the M8 library were screened by plaque hybridization using the ^{32}P-labeled rat IRS-1 cDNA 5.4-kb fragment as a probe (7). Hybridization was performed in 50% formamide, 0.5% SDS, 50 mM EDTA, 3 × SSC (1 × SSC = 150 mM NaCl/15 mM sodium citrate), 1 × Denhardt's solution, and 100 μg/ml salmon sperm DNA at 42°C. Filters were washed with 2 × SSC and 0.1% SDS at room temperature for 15 min four times and with 0.2 × SSC and 0.1% SDS at 52°C for 40 min. After three rounds of screening, λE1 and λE2 (corresponding to nucleotides 1257–2303 in Fig. 1*A*) were obtained from the M8 library. λE3, λE4, λE5, λE6, λE8, λE14, and λE27 were obtained from the M12 library using the rat liver IRS-1 cDNA and the insert of λE1 as probes. λE3, λE4, and λE27 were identical and corresponded to nucleotide 4128 to 3′ end in Fig. 1*A*; λE5 to λE8 correspond to nucleotide 3672, 5321, 4860, 5247 to the 3′ end, respectively. λE36, which covered the rest of the cDNA (nucleotide 1–4104 in Fig. 1*A*), was obtained from the fetal skeletal muscle library using the rat liver IRS-1 cDNA and inserts of λE1 and λE5 as probes. The Bluescript SK$^-$ plasmids containing the cDNA insert were released from the λ*Zap*II vector by in vivo excision with the helper phage R408 as described in the manufacturer's instructions (Stratagene). The insert of λE36 was digested, purified, and subcloned in the *Eco*RI site of pBluescript KSII (Stratagene). The nucleotide sequence of the cDNA was determined by the dideoxy chain-termination method (20) using the Sequenase DNA sequencing kit (U.S. Biochemical, Cleveland, OH). Both strands of the clones were sequenced using T3 and T7 primers, as well as using synthetic oligonucleotide primers deduced from the partially determined sequence.

Cloning the human IRS-1 gene. A human genomic library in EMBL 3A was screened using the inserts of λE1 and λE5. Two clones, λL9 and λL10, were obtained after screening of 1×10^6 placques. Inserts of these clones were digested with *Eco*RI, subcloned in the *Eco*RI site of pBluescript KSII, and selected by colony hybridization method (6). The clones were further analyzed by polymerase chain reaction and sequencing with the primers deduced from the human IRS-1 cDNA.

Expression of human IRS-1 in CHO cells. A 4.1-kb fragment of human IRS-1 cDNA (943-5027 in Fig. 1*A*) was inserted in the expression vector pCMV-His (22) using standard molecular biological techniques (23), and named pCMV-His-hIRS-1. CHO cells or CHO/hIR cells (6) were transfected with 10 μg of pCMV-His-hIRS-1 using the calcium phosphate method (24) and selected with 10 mM histidinol (Sigma, St. Louis, MO). Surviving clones were picked as described previously (25), and clones expressing human IRS-1 were selected by Western blotting as described below.

Western blot analysis. CHO cells or CHO/hIR cells that express high levels of human IRS-1 (called CHO/hIRS-1, CHO/hIR/hIRS-1, respectively) or control cells transfected with only pCMV-His (CHO/His, CHO/hIR/His, respectively) were cultured in 10-cm dishes (Nunc, Roskilde, Denmark) in F12 medium with 10% FBS (Sigma) at 37°C in a 5% CO_2 incubator. For Western blot experiments, 90% confluent cultures were serum deprived for 14–16 h before the study. Insulin (final concentration 10^{-7} M) was added, and the incubation was continued at 37°C for 1 min. The cell monolayer was frozen with liquid nitrogen, thawed, and homogenized with 0.5 ml of the extraction buffer (100 mM Tris, pH 8.0, containing 100 mM sodium fluoride, 4 mM EDTA, 2 mM sodium vanadate, 3.4 mg/ml PMSF, 100 μg/ml aprotinin, and 1% Triton X-100), and insoluble materials were sedimented by microcentrifugation at 13,000 *g* for 15 min. The supernatant was dissolved in Laemmli sample buffer with 100 mM DTT at 100°C for 5 min, and 80 μg of protein from each cell extract was subjected to SDS-PAGE (6% bis-acrylamide) in a Bio-Rad miniature slab

gel apparatus and transferred to nitrocellulose (Schleicher and Schuell, Keene, NH) or PVDF membranes (Millipore, Bedford, MA) for 1 h at 90 V in the Bio-Rad miniature transfer apparatus as described by Towbin et al. (26), except that 0.02% SDS was added to the transfer buffer to enhance elution of high molecular mass proteins. The blots were incubated with the appropriate antibodies, washed, and incubated with [^{125}I]protein A (ICN, Irvine, CA), then autoradiography was performed as described previously (13).

Northern blot analysis. Poly(A)$^+$ RNA (2 μg) purified from human adult skeletal muscle was fractionated on a 1% agarose/formaldehyde gel, and transferred onto a nitrocellulose membrane in a 20 × SSC for 20 h and fixed to the membrane by baking at 80°C for 2 h. A 4-kb cDNA fragment of human IRS-1 containing the entire coding region (943-4836 in Fig. 1*A*) was purified on an agarose gel, ^{32}P-labeled using a random-primed oligolabeling kit (Pharmacia) and used for hybridization (1 × 10^6 cpm/ml) as described above. Also, a premade Northern blot (Human Multiple Tissue Northern Blot, Clontech, Palo Alto, CA) that contains 2 μg of poly(A)$^+$ RNA from each tissue was used for Northern blot analysis using the same probe (1 × 10^6 cpm/ml) as described in the manufacturer's instructions.

Quantitative PCR. Primers for reverse transcription and the PCR are designed from the cDNA sequences. For human IRS-1, the 5′ primer was AGTGGCCATGGCT CCAC (corresponding to nucleotides 2212–2228 in Fig. 1*A*) and the 3′ primer was TTGCCACCCATGCAGAT (nucleotides 2416–2432); for the hIR, the 5′ primer was TTCCGAGACCTCAGTTTC (corresponding to nucleotides 355–372 in a previous study [3]) and the 3′ primer was TGTGACTTACAGATGGT (nucleotides 799–815); for human β_2 microglobulin, the 5′ primer was TTCAGCA AGGACTGGTCT (corresponding to nucleotides 1498–1516 in a previous study [27]) and the 3′ primer was CTGCTTACATGTCTCGAT (nucleotides 2235–2252). Primers for the hIR and β_2 microglobulin spanned at least one intron to avoid contamination of the PCR product by amplification of genomic DNA.

One microgram of total RNA from second trimester human fetal tissue was reverse transcribed into cDNA using all three 3′ primers (1 pmol primer/reaction). After [^{32}P]end-labeling (28) of the PCR primers, 5% of the total cDNA was used as a template for the PCR reaction. The final PCR conditions were 10 pmol of each primer/reaction, denaturing at 94°C for 30 s, annealing at 57°C for 30 s, and extension at 72°C for 1 min. Based on pilot tests, linear amplification of IRS-1 mRNA was achieved using 50 ng of total RNA/reaction and 21–27 cycles (data not shown). The final PCR products were separated on a 5% acrylamide gel in 0.5 × TBE buffer. Gels were dried, subjected to autoradiography, and the radioactivity in the bands was determined by Cerenkov counting. As a negative control, the reverse transcription reaction followed by PCR was performed without addition of RNA. As positive controls, ~10 ng of human IRS-1 or insulin receptor cDNA was used as a PCR template. To determine the amount of IRS-1 PCR product derived from any contaminating genomic DNA, the reverse transcription and PCR reactions were performed by omitting the reverse transcriptase.

Southern blot analysis using high or moderate stringency. Genomic DNA strands (20 μg), prepared from human leukocytes, were digested with restriction enzymes *Bam*HI, *Eco*RI, *Hind*III, and *Kpn*I, divided into two samples, separated on a 1% agarose gel, transferred onto a nitrocellulose filter, and fixed to the filter by baking at 80°C for 2 h. The filter was cut into two pieces and hybridized with the ^{32}P-labeled 1-kb human IRS-1 cDNA fragment located in the coding region (1257–2303 in Fig. 1*A*). For high stringency analysis, the conditions were identical to those for the Northern blot, except that the temperature of the second wash was 57°C. For moderate stringency analysis, the filter was hybridized with the same prehybridization buffer except that the concentration of formamide was reduced to 40% and the temperature reduced to 35°C. These filters were washed with 2 × SSC and 0.1% SDS at 35°C for 15 min, 0.5 × SSC and 0.1% SDS at 40°C for 15 min, and 0.1 × SSC and 0.1% SDS at 50°C for 30 min. Both filters were then subjected to autoradiography.

Chromosomal localization of IRS-1 gene by in situ hybridization. The cDNA probe was nick-translated with [^{3}H]dATP and [^{3}H]dCTP to a specific activity of 2.45 × 10^7 cpm/μg and used for hybridization at a concentration of 25 ng/ml. Hybridization to human chromosome spreads, post hybridization wash, emulsion autoradiography, and silver grain analysis were conducted as described previously (29).

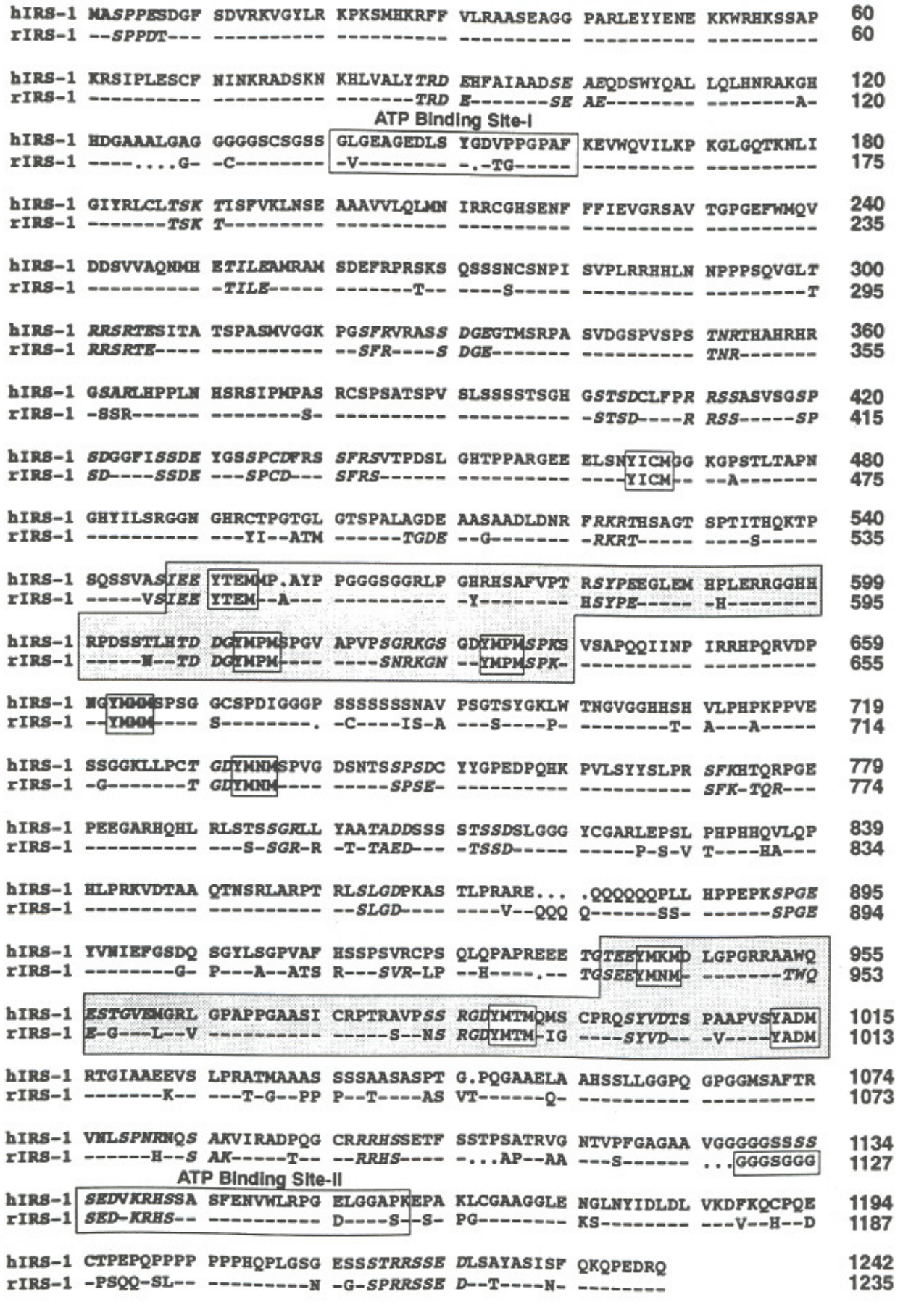
RESULTS

Cloning of the human muscle IRS-1 cDNA. The full-length cDNA for human IRS-1 was sequenced and assembled from a total of 10 fragments obtained from three different human muscle cDNA libraries as described in METHODS. Sequencing of all of these clones and construction of the complete human IRS-1 cDNA revealed that human IRS-1 has a very high degree of homology with rat liver IRS-1. Human IRS-1 cDNA contains a 3726-bp open reading frame starting at a Kozak consensus translation start site (30,31) at base 1021 and ending at base 4746 with a TAG (Fig. 1*A*). The nucleotide sequences differed between clones at 2 sites, both of which would result in coding different amino acids (T in λE1 and λE2 and C in λE36 at nucleotide 1346; T in λE5 and A in λE36 at nucleotide 3677 in Fig. 1*A*). Because the clone λE5 was derived from the NIDDM patient, it is possible that one of the amino acid changes from His^{886} to Leu^{886} has some biological significance. Further analysis of this point is necessary.

The 5′-untranslated region of human IRS-1 cDNA is 432 bp longer than that of rat IRS-1 cDNA, and the region overlapping both cDNAs has 73% identity at the nucleotide level. The 5′-untranslated region has one TC-rich region (TCCC repetitive sequence) and 4 potential binding sites for the transcription factor Sp1 (32,33). At least one of these Sp1-binding sites (nucleotides 442–447 in Fig. 1*A*) is conserved in both the rat and human IRS-1 cDNAs. Analysis of the promoter region of the mouse IRS-1 gene has revealed several potential transcription

```
   1: CGGCGGCGCG GTCGGAGGGG GCCGGCGCGC AGAGCCAGAC GCCGCCGCTT GTTTTGGTTG GGGCTCTCGG CAACTCTCCG AGGAGGAGGA GGAGGAGGGA GGAGGGGAGA

 111:  AGTAACTGCA GCGGCAGCGC CCTCCCGAGG AACAGGCGTC TTCCCCGAAC CCTTCCCAAA CCTCCCCCAT CCCCTCTCGC CCTTGTCCCC TCCCCTCCTC CCCAGCCGC

 220: C TGGAGCGAGG GGCAGGGATG AGTCTGTCCC TCCGGCCGGT CCCCAGCTGC AGTGGCTGCC CGGTATCGTT TCGCATGGAA AAGCCACTTT CTCCACCCGC CGAGATGG
                                                                                                                          *
 329: GC CCGGATGGGG CTGCAGAGGA CGCGCCCGCG GGCGGCGGCA GCAGCAGCAG CAGCAGCAGC AGCAACAGCA ACAGCCGCAG CGCCGCGGTC TCTGCGACTG AGCTGGT

 438: ATT TGGGCGGCTG GTGGCGGCTG GGACGGTTGG GGGGTGGGAG GAGGCGAAGG AGGAGGGAGA ACCCCGTGCA ACGTTGGGAC TTGGCAACCC GCCTCCCCCT GCCCAA

 547: GGAT ATTTAATTTG CCTCGGGAAT CGCTGCTTCC AGAGGGGAAC TCAGGAGGGA AGGCGCGCGC GCGCGCGCGC TCCTGGAGGG GCACCGCAGG GACCCCCGAC TGTCG

 656: CCTCC CTGTGCCGGA CTCCAGCCGG GGCGACGAGA GATGCATCTT CGCTCCTTCC TGGTGGCGGC GGCGGCTGAG AGGAGACTTG GCTCTCGGAG GATCGGGGCT GCCC

 765: TCACCC CGGACGCACT GCCTCCCCGC CGGCGTGAAG CGCCCGAAAA CTCCGGTCGG GCTCTCTCCT GGGCTCAGCA GCTGCGTCCT CCTTCAGCTG CCCCTCCCCG GCG

 874: CGGGGGG CGGCGTGGAT TTCAGAGTCG GGGTTTCTGC TGCCTCCAGC CCTGTTTGCA TGTGCCGGGC CGCGGCGAGG AGCCTCCGCC CCCCACCCGG TTGTTTTTCG GA

 983: GCCTCCCT CTGCTCAGCG TTGGTGGTGG CGGTGGCAGC ATG GCG AGC CCT CCG GAG AGC GAT GGC TTC TCG GAC GTG CGC AAG GTG GGC TAC CTG CG
   1;                                      met ala ser pro pro glu ser asp gly phe ser asp val arg lys val gly tyr leu ar

1080: C AAA CCC AAG AGC ATG CAC AAA CGC TTC TTC GTA CTG CGC GCG GCC AGC GAG GCT GGG GGC CCG GCG CGC CTC GAG TAC TAC GAG AAC GA
  21: g lys pro lys ser met his lys arg phe phe val leu arg ala ala ser glu ala gly gly pro ala arg leu glu tyr tyr glu asn gl

1170: G AAG AAG TGG CGG CAC AAG TCG AGC GCC CCC AAA CGC TCG ATC CCC CTT GAG AGC TGC TTC AAC ATC AAC AAG CGG GCT GAC TCC AAG AA
  51: u lys lys trp arg his lys ser ser ala pro lys arg ser ile pro leu glu ser cys phe asn ile asn lys arg ala asp ser lys as
                                                                                                                        T
1260: C AAG CAC CTG GTG GCT CTC TAC ACC CGG GAC GAG CAC TTT GCC ATC GCG GCG GAC AGC GAG GCC GAG CAA GAC AGC TGG TAC CAG GCT CT
  81: n lys his leu val ala leu tyr thr arg asp glu his phe ala ile ala ala asp ser glu ala glu gln asp ser trp tyr gln ala le
                                                                                                                        val
1350: C CTA CAG CTG CAC AAC CGT GCT AAG GGC CAC CAC GAC GGA GCT GCG GCC CTC GGG GCG GGA GGT GGT GGG GGC AGC TGC AGC GGC AGC TC
 111: u leu gln leu his asn arg ala lys gly his his asp gly ala ala ala leu gly ala gly gly gly gly gly ser cys ser gly ser se

1440: C GGC CTT GGT GAG GCT GGG GAG GAC TTG AGC TAC GGT GAC GTG CCC CCA GGA CCC GCA TTC AAA GAG GTC TGG CAA GTG ATC CTG AAG CC
 141: r gly leu gly glu ala gly glu asp leu ser tyr gly asp val pro pro gly pro ala phe lys glu val trp gln val ile leu lys pr

1530: C AAG GGC CTG GGT CAG ACA AAG AAC CTG ATT GGT ATC TAC CGC CTT TGC CTG ACC AGC AAG ACC ATC AGC TTC GTG AAG CTG AAC TCG GA
 171: o lys gly leu gly gln thr lys asn leu ile gly ile tyr arg leu cys leu thr ser lys thr ile ser phe val lys leu asn ser gl

1620: G GCA GCG GCC GTG GTG CTG CAG CTG ATG AAC ATC AGG CGC TGT GGC CAC TCG GAA AAC TTC TTC TTC ATC GAG GTG GGC CGT TCT GCC GT
 201: u ala ala ala val val leu gln leu met asn ile arg arg cys gly his ser glu asn phe phe phe ile glu val gly arg ser ala va

1710: G ACG GGG CCC GGG GAG TTC TGG ATG CAG GTG GAT GAC TCT GTG GTG GCC CAG AAC ATG CAC GAG ACC ATC CTG GAG GCC ATG CGG GCC AT
 231: l thr gly pro gly glu phe trp met gln val asp asp ser val val ala gln asn met his glu thr ile leu glu ala met arg ala me

1800: G AGT GAT GAG TTC CGC CCT CGC AGC AAG AGC CAG TCC TCG TCC AAC TGC TCT AAC CCC ATC AGC GTC CCC CTG CGC CGG CAC CAT CTC AA
 261: t ser asp glu phe arg pro arg ser lys ser gln ser ser ser asn cys ser asn pro ile ser val pro leu arg arg his his leu as

1890: C AAT CCC CCG CCC AGC CAG GTG GGG CTG ACC CGC CGA TCA CGC ACT GAG AGC ATC ACC GCC ACC TCC CCG GCC AGC ATG GTG GGC GGG AA
 291: n asn pro pro pro ser gln val gly leu thr arg arg ser arg thr glu ser ile thr ala thr ser pro ala ser met val gly gly ly

1980: G CCA GGC TCC TTC CGT GTC CGC GCC TCC AGT GAC GGC GAA GGC ACC ATG TCC CGC CCA GCC TCG GTG GAC GGC AGC CCT GTG AGT CCC AG
 321: s pro gly ser phe arg val arg ala ser ser asp gly glu gly thr met ser arg pro ala ser val asp gly ser pro val ser pro se

2070: C ACC AAC AGA ACC CAC GCC CAC CGG CAT CGG GGC AGC GCC CGG CTG CAC CCC CCG CTC AAC CAC AGC CGC TCC ATC CCC ATG CCG GCT TC
 351: r thr asn arg thr his ala his arg his arg gly ser ala arg leu his pro pro leu asn his ser arg ser ile pro met pro ala se

2160: C CGC TGC TCG CCT TCG GCC ACC AGC CCG GTC AGT CTG TCG TCC AGT AGC ACC AGT GGC CAT GGC TCC ACC TCG GAT TGT CTC TTC CCA CG
 381: r arg cys ser pro ser ala thr ser pro val ser leu ser ser ser ser thr ser gly his gly ser thr ser asp cys leu phe pro ar

2250: G CGA TCT AGT GCT TCG GTG TCT GGT TCC CCC AGC GAT GGC GGT TTC ATC TCC TCG GAT GAG TAT GGC TCC AGT CCC TGC GAT TTC CGG AG
 411: g arg ser ser ala ser val ser gly ser pro ser asp gly gly phe ile ser ser asp glu tyr gly ser ser pro cys asp phe arg se

2340: T TCC TTC CGC AGT GTC ACT CCG GAT TCC CTG GGC CAC ACC CCA CCA GCC CGC GGT GAG GAG GAG CTA AGC AAC TAT ATC TGC ATG GGT GG
 441: r ser phe arg ser val thr pro asp ser leu gly his thr pro pro ala arg gly glu glu glu leu ser asn tyr ile cys met gly gl

2430: C AAG GGG CCC TCC ACC CTG ACC GCC CCC AAC GGT CAC TAC ATT TTG TCT CGG GGT GGC AAT GGC CAC CGC TGC ACC CCA GGA ACA GGC TT
 471: y lys gly pro ser thr leu thr ala pro asn gly his tyr ile leu ser arg gly gly asn gly his arg cys thr pro gly thr gly le

2520: G GGC ACG AGT CCA GCC TTG GCT GGG GAT GAA GCA GCC AGT GCT GCA GAT CTG GAT AAT CGG TTC CGA AAG AGA ACT CAC TCG GCA GGC AC
 501: u gly thr ser pro ala leu ala gly asp glu ala ala ser ala ala asp leu asp asn arg phe arg lys arg thr his ser ala gly th
```

```
2610: A TCC CCT ACC ATT ACC CAC CAG AAG ACC CCG TCC CAG TCC TCA GTG GCT TCC ATT GAG GAG TAC ACA GAG ATG ATG CCT GCC TAC CCA CC
 531: r ser pro thr ile thr his gln lys thr pro ser gln ser ser val ala ser ile glu glu tyr thr glu met met pro ala tyr pro pr

2700: A GGA GGT GGC AGT GGA GGC CGA CTG CCG GGA CAC AGG CAC TCC GCC TTC GTG CCC ACC CGC TCC TAC CCA GAG GAG GGT CTG GAA ATG CA
 561: o gly gly gly ser gly gly arg leu pro gly his arg his ser ala phe val pro thr arg ser tyr pro glu glu gly leu glu met hi

2790: C CCC TTG GAG CGT CGG GGG GGG CAC CAC CGC CCA GAC AGC TCC ACC CTC CAC ACG GAT GAT GGC TAC ATG CCC ATG TCC CCA GGG GTG GC
 591: s pro leu glu arg arg gly gly his his arg pro asp ser ser thr leu his thr asp asp gly tyr met pro met ser pro gly val al

2880: C CCA GTG CCC AGT GGC CGA AAG GGC AGT GGA GAC TAT ATG CCC ATG AGC CCC AAG AGC GTA TCT GCC CCA CAG CAG ATC ATC AAT CCC AT
 621: a pro val pro ser gly arg lys gly ser gly asp tyr met pro met ser pro lys ser val ser ala pro gln gln ile ile asn pro il

2970: C AGA CGC CAT CCC CAG AGA GTG GAC CCC AAT GGC TAC ATG ATG ATG TCC CCC AGC GGT GGC TGC TCT CCT GAC ATT GGA GGT GGC CCC AG
 651: e arg arg his pro gln arg val asp pro asn gly tyr met met met ser pro ser gly gly cys ser pro asp ile gly gly gly pro se

3060: C AGC AGC AGC AGC AGC AGC AAC GCC GTC CCT TCC GGG ACC AGC TAT GGA AAG CTG TGG ACA AAC GGG GTA GGG GGC CAC CAC TCT CAT GT
 681: r ser ser ser ser ser ser asn ala val pro ser gly thr ser tyr gly lys leu trp thr asn gly val gly gly his his ser his va

3150: C TTG CCT CAC CCC AAA CCC CCA GTG GAG AGC AGC GGT GGT AAG CTC TTA CCT TGC ACA GGT GAC TAC ATG AAC ATG TCA CCA GTG GGG GA
 711: l leu pro his pro lys pro pro val glu ser ser gly gly lys leu leu pro cys thr gly asp tyr met asn met ser pro val gly as

3240: C TCC AAC ACC AGC AGC CCC TCC GAC TGC TAC TAC GGC CCT GAG GAC CCC CAG CAC AAG CCA GTC CTC TCC TAC TAC TCA TTG CCA AGA TC
 741: p ser asn thr ser ser pro ser asp cys tyr tyr gly pro glu asp pro gln his lys pro val leu ser tyr tyr ser leu pro arg se

3330: C TTT AAG CAC ACC CAG CGC CCC GGG GAG CCG GAG GAG GGT GCC CGG CAT CAG CAC CTC CGC CTT TCC ACT AGC TCT GGT CGC CTT CTC TA
 771: r phe lys his thr gln arg pro gly glu pro glu glu gly ala arg his gln his leu arg leu ser thr ser ser gly arg leu leu ty

3420: T GCT GCA ACA GCA GAT GAT TCT TCC TCT TCC ACC AGC AGC GAC AGC CTG GGT GGG GGA TAC TGC GGG GCT AGG CTG GAG CCC AGC CTT CC
 801: r ala ala thr ala asp asp ser ser ser ser thr ser ser asp ser leu gly gly gly tyr cys gly ala arg leu glu pro ser leu pr

3510: A CAT CCC CAC CAT CAG GTT CTG CAG CCC CAT CTG CCT CGA AAG GTG GAC ACA GCT GCT CAG ACC AAT AGC CGC CTG GCC CGG CCC ACG AG
 831: o his pro his his gln val leu gln pro his leu pro arg lys val asp thr ala ala gln thr asn ser arg leu ala arg pro thr ar
                                                                                                                  T
3600: G CTG TCC CTG GGG GAT CCC AAG GCC AGC ACC TTA CCT CGG GCC CGA GAG CAG CAG CAG CAG CAG CAG CCC TTG CTG CAC CCT CCA GAG CC
 861: g leu ser leu gly asp pro lys ala ser thr leu pro arg ala arg glu gln gln gln gln gln gln pro leu leu his pro pro glu pr
                                                                                                                 leu
3690: C AAG AGC CCG GGG GAA TAT GTC AAT ATT GAA TTT GGG AGT GAT CAG TCT GGC TAC TTG TCT GGC CCG GTG GCT TTC CAC AGC TCA CCT TC
 891: o lys ser pro gly glu tyr val asn ile glu phe gly ser asp gln ser gly tyr leu ser gly pro val ala phe his ser ser pro se

3780: T GTC AGG TGT CCA TCC CAG CTC CAG CCA GCT CCC AGA GAG GAA GAG ACT GGC ACT GAG GAG TAC ATG AAG ATG GAC CTG GGG CCG GGC CG
 921: r val arg cys pro ser gln leu gln pro ala pro arg glu glu glu thr gly thr glu glu tyr met lys met asp leu gly pro gly ar

3870: G AGG GCA GCC TGG CAG GAG AGC ACT GGG GTC GAG ATG GGC AGA CTG GGC CCT GCA CCT CCC GGG GCT GCT AGC ATT TGC AGG CCT ACC CG
 951: g arg ala ala trp gln glu ser thr gly val glu met gly arg leu gly pro ala pro pro gly ala ala ser ile cys arg pro thr ar

3960: G GCA GTG CCC AGC AGC CGG GGT GAC TAC ATG ACC ATG CAG ATG AGT TGT CCC CGT CAG AGC TAC GTG GAC ACC TCG CCA GCT GCC CCT GT
 981: g ala val pro ser ser arg gly asp tyr met thr met gln met ser cys pro arg gln ser tyr val asp thr ser pro ala ala pro va

4050: A AGC TAT GCT GAC ATG CGA ACA GGC ATT GCT GCA GAG GAG GTG AGC CTG CCC AGG GCC ACC ATG GCT GCT GCC TCC TCA TCC TCA GCA GC
1011: l ser tyr ala asp met arg thr gly ile ala ala glu glu val ser leu pro arg ala thr met ala ala ala ser ser ser ser ala al

4140: C TCT GCT TCC CCG ACT GGG CCT CAA GGG GCA GCA GAG CTG GCT GCC CAC TCG TCC CTG CTG GGG GGC CCA CAA GGA CCT GGG GGC ATG AG
1041: a ser ala ser pro thr gly pro gln gly ala ala glu leu ala ala his ser ser leu leu gly gly pro gln gly pro gly gly met se

4230: C GCC TTC ACC CGG GTG AAC CTC AGT CCT AAC CGC AAC CAG AGT GCC AAA GTG ATC CGT GCA GAC CCA CAA GGG TGC CGG CGG AGG CAT AG
1071: r ala phe thr arg val asn leu ser pro asn arg asn gln ser ala lys val ile arg ala asp pro gln gly cys arg arg arg his se

4320: C TCC GAG ACT TTC TCC TCA ACA CCC AGT GCC ACC CGG GTG GGC AAC ACA GTG CCC TTT GGA GCG GGG GCA GCA GTA GGG GGC GGT GGC GG
1101: r ser glu thr phe ser ser thr pro ser ala thr arg val gly asn thr val pro phe gly ala gly ala ala val gly gly gly gly gl

4410: T AGC AGC AGC AGC AGC GAG GAT GTG AAA CGC CAC AGC TCT GCT TCC TTT GAG AAT GTG TGG CTG AGG CCT GGG GAG CTT GGG GGA GCC CC
1131: y ser ser ser ser ser glu asp val lys arg his ser ser ala ser phe glu asn val trp leu arg pro gly glu leu gly gly ala pr

4500: C AAG GAG CCA GCC AAA CTG TGT GGG GCT GCT GGG GGT TTG GAG AAT GGT CTT AAC TAC ATA GAC CTG GAT TTG GTC AAG GAC TTC AAA CA
1161: o lys glu pro ala lys leu cys gly ala ala gly gly leu glu asn gly leu asn tyr ile asp leu asp leu val lys asp phe lys gl

4590: G TGC CCT CAG GAG TGC ACC CCT GAA CCG CAG CCT CCC CCA CCC CCA CCC CCT CAT CAA CCC CTG GGC AGC GGT GAG AGC AGC TCC ACC CG
1191: n cys pro gln glu cys thr pro glu pro gln pro pro pro pro pro pro pro his gln pro leu gly ser gly glu ser ser ser thr ar

4680: C CGC TCA AGT GAG GAT TTA AGC GCC TAT GCC AGC ATC AGT TTC CAG AAG CAG CCA GAG GAC CGT CAG TAG C TCAACTGGAC ATCACAGCAG AA
1221: g arg ser ser glu asp leu ser ala tyr ala ser ile ser phe gln lys gln pro glu asp arg gln ***                        †
```


ATP Binding Site-I
ATP Binding Site-II

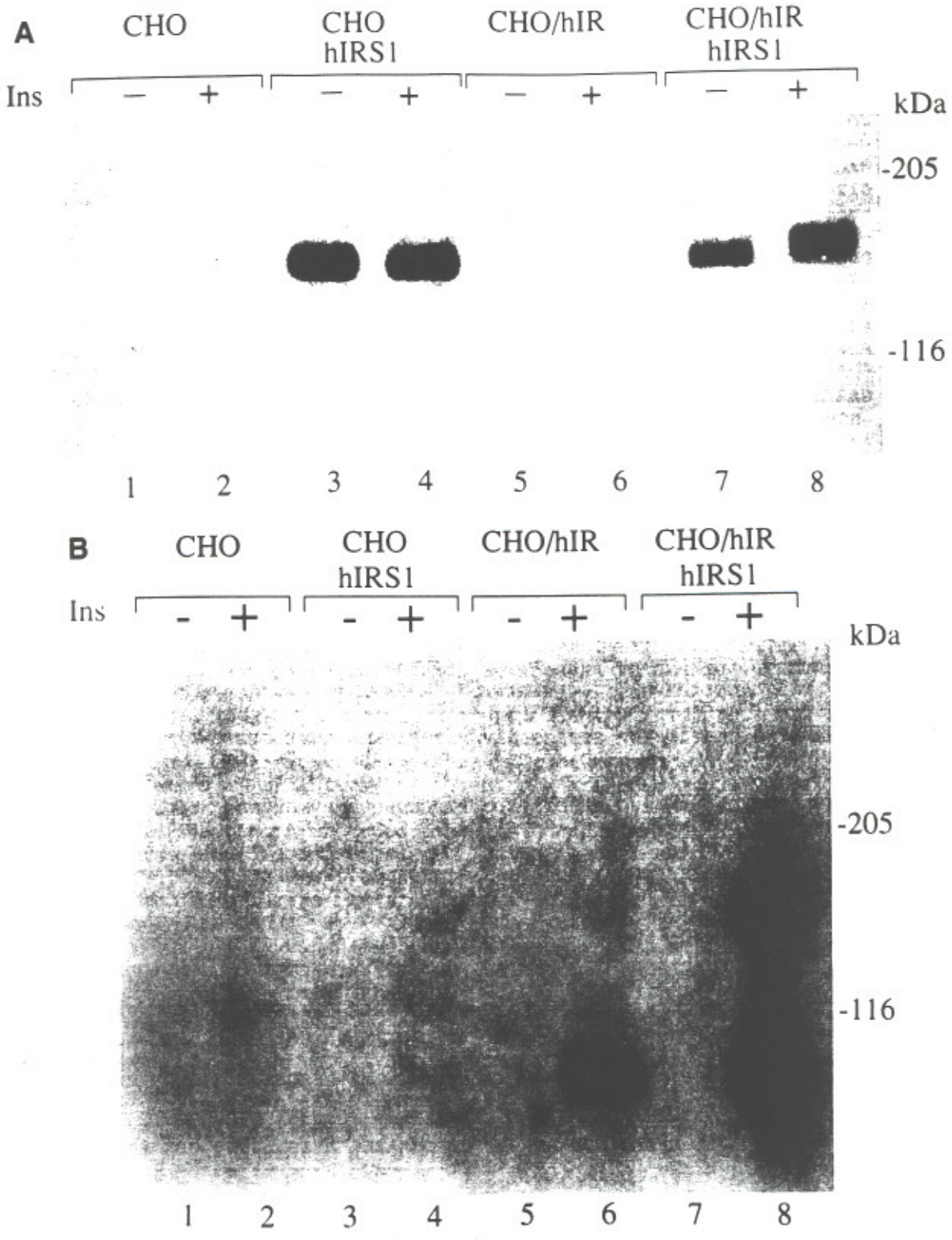
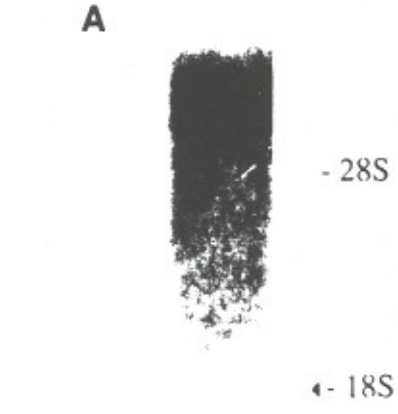


FIG. 2. Expression and characterization of human IRS-1 in CHO cells or CHO/hIR cells. CHO and CHO/hIRS-1 represent the CHO cells that carry pCMV-His and pCMV-His-hIRS-1, respectively, and CHO/hIR and CHO/hIR/hIRS-1 represent the CHO/hIR cells that carry pCMV-His and pCMV-His-hIRS-1, respectively. These cells were treated (lanes 2, 4, 6, and 8) or untreated (lanes 1, 3, 5, and 7) with insulin (10^{-7} M) and extracted; and equal aliquots (80 μg of protein) were subjected to SDS-PAGE followed by Western blot analysis using the α-hIRS-1 antibody (*A*), or antiphosphotyrosine antibody, αPY (*B*).



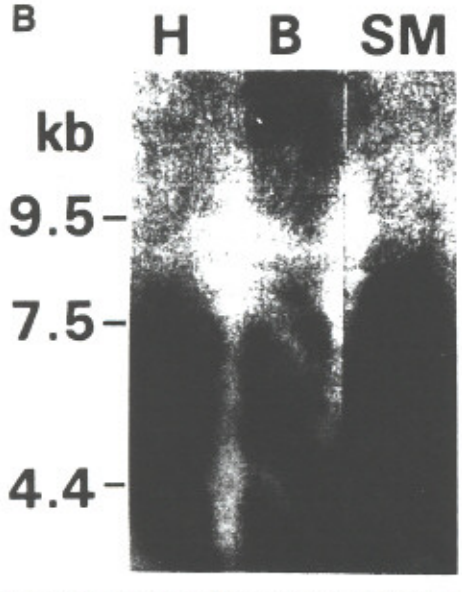


FIG. 3. Northern blot analysis of human skeletal muscle IRS-1 mRNA. Nitrocellulose membrane on which 2 μg of poly(A)$^+$ RNA prepared from human skeletal muscle was transferred (*A*) or human multiple tissue Northern blot (Clontech) (*B*) were subjected to Northern blot analysis. (*H*), heart; (*B*), brain; (*SM*), skeletal muscle. The films were exposed for 3 days at −70°C.

terminal ATP-binding site, including the two GXGXXG motifs, is conserved, but both of the COOH-terminal sites are not, because the third Gly of these motifs has been replaced (one with Val and one with Ser). In all Tyr kinases and most Ser kinases whose sequences are known, the third Gly is conserved. However, in a few Ser kinases, this Gly is substituted for by either Ser or Ala (37).

Expression and characterization of human IRS-1 in CHO cell and CHO/hIR cells. To confirm that the clone was human IRS-1, a 4.1-kb cDNA fragment corresponding to the entire coding region was inserted in the multicloning site of pCMV-His and transfected into CHO cells or CHO cells overexpressing the hIR (CHO/hIR). Approximately 50 clones were obtained in each transfection after histidinol selection, and 12 clones were examined for human IRS-1 expression by Western blot using a polyclonal antibody raised to baculovirus-expressed rat IRS-1 (α-bIRS-1). Three clones expressing high levels of human IRS-1 were obtained in CHO cells, and two were obtained in CHO/hIR cells; these were termed CHO/hIRS-1 (clones A10, B3, and B22) and CHO/hIR/hIRS-1 (clones 26 and 48), respectively.

Both CHO/hIRS-1 and CHO/hIR/hIRS-1 cells expressed human IRS-1 as detected by Western blot analysis with anti–COOH-terminal (α-CT-IRS-1) and anti-baculovirus IRS-1 (α-bIRS-1) antibodies. IRS-1 migrated as a protein of ~180,000 M_r (Fig. 2*A*, lanes 3, 4, and 7) despite a calculated M_r of 132,000. The apparent M_r increased after insulin stimulation, especially in CHO/hIR/hIRS-1 cells (Fig. 2*A*, lane 8), probably because of Tyr phosphorylation by the insulin receptor. A low level of endogenous IRS-1 could also be observed in both CHO/His and CHO/hIR/His cells (Fig. 3*A*, lanes 1, 2, 5, and 6) upon longer exposure of autoradiographs using the α-bIRS-1 antibody, but could not be seen in blots with the α-CT-IRS-1 antibody (data not shown).

The amount of IRS-1 expressed in CHO/hIRS-1 cells was comparable or even slightly higher than that of CHO/hIR/hIRS-1; however, the magnitude of Tyr phosphorylation of IRS-1 after insulin stimulation was completely different in these two types of cells (Fig. 2*B*). We observed a 10-fold higher insulin-stimulated Tyr phosphorylation (measured by densitometry) in CHO/IR/IRS-1 cells than in cells expressing human IRS-1 alone. A slight increase (1.8-fold) of IRS-1 Tyr phosphorylation was observed in CHO/hIRS-1 cells after insulin stimulation compared with CHO/His cells (lanes 2 and 4); however, in the absence of overexpression of the insulin receptor, Tyr phosphorylation of IRS-1 in CHO/hIRS-1 cells was still less than that of CHO/hIR cells (0.6-fold) (lanes 4 and 6). The largest increase of IRS-1 Tyr phosphorylation after insulin stimulation was seen in CHO/hIR/hIRS-1 cells that were overexpressing both the hIR and human IRS-1. These data indicate that *1*) cloned human IRS-1 cDNA is correctly expressed in those cells; *2*) human IRS-1 is Tyr phosphorylated by insulin stimulation in vivo; and *3*) Tyr phosphorylation of IRS-1 is regulated by both the level of expression of IRS-1 and, even more, by the level of insulin receptor Tyr kinase.

Northern blot analysis and expression of human IRS-1 in human tissues. Northern blot analysis revealed that the abundance of human IRS-1 mRNA was low with no hybridization signals detected at either 10 or 50 μg of total RNA prepared from human skeletal muscle (data not shown). Using a 4-kb human IRS-1 cDNA probe

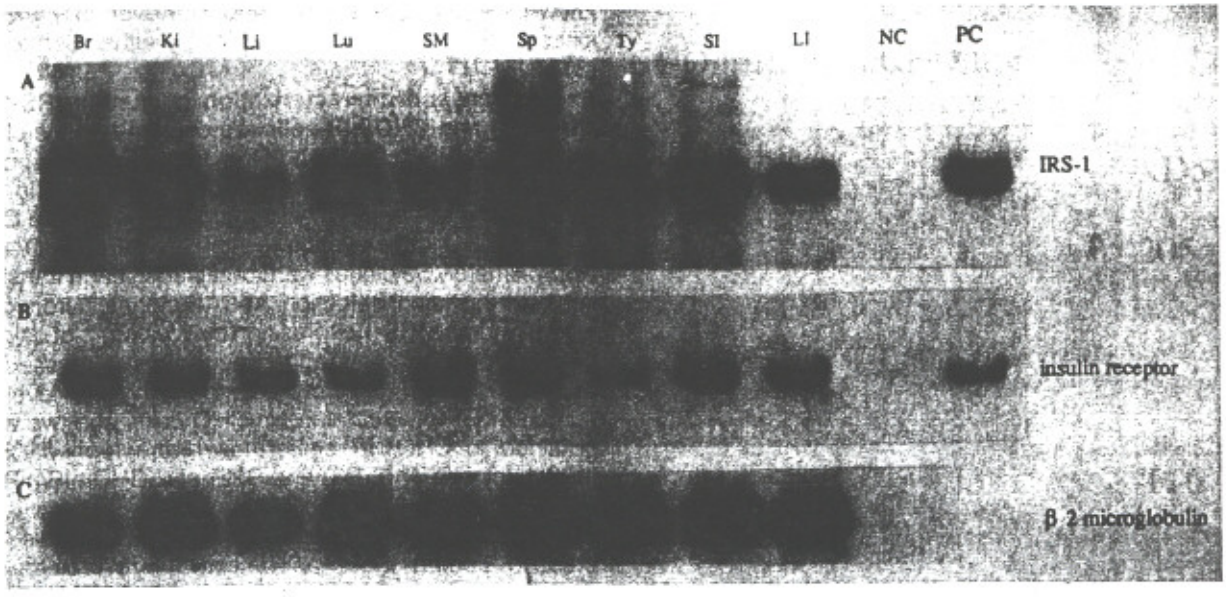


FIG. 4. Quantitation of IRS-1 mRNA in human fetal tissues by quantitative PCR. One microgram of total RNA prepared from human fetal tissues—(*Br*), brain; (*Ki*), kidney; (*Li*), liver; (*Lu*), lung; (*SM*), skeletal muscle; (*Sp*), spleen; (*Ty*), thymus; (*SI*), small intestine; and (*LI*), large intestine—was reverse transcribed into cDNA in the presence of the three 3′ primers (1 pmol of each primer/reaction). One twentieth of total cDNA was subjected to PCR using end-labeled primers of human IRS-1 (*A*), hIR (*B*), or β_2 microglobulin (*C*). Reactions without the RNA sample were performed as negative controls (lane *NC*). As positive controls, ~10 ng of each cDNA was used as a template of PCR for human IRS-1 and hIR (lane *PC*).

containing the entire coding region, two mRNAs of 6.9 and 6 kb could be detected in poly(A)$^+$ RNA isolated from adult human skeletal muscle, heart, and brain (Figs. 3*A* and *B*). To examine the expression of IRS-1 mRNA in human fetal tissues and to compare its expression with that of insulin receptor mRNA, semiquantitative PCR using total RNA from human fetal tissues was performed. As a positive control between the tissues, β_2 microglobin, which is known to be expressed in every tissue, was used. By this method, IRS-1 and insulin receptor mRNAs were detected in all tissues examined (Fig. 4). IRS-1 mRNA was most abundant in the small intestine, brain, and kidney, and showed relatively lower abundance in other tissues. Insulin receptor mRNA was abundant in the small intestine, brain, kidney, spleen, and large intestine, and was distributed more evenly among tissues when compared with IRS-1 mRNA. In control PCR reactions using samples not subjected previously to reverse transcription, the level of IRS-1 PCR product derived from contaminating genomic DNA was very small and almost equal in each reaction (data not shown).

Southern blot analysis using high or moderate stringency conditions. To determine the potential number of copies of the IRS-1 gene or related genes, a 1-kb fragment of the human IRS-1 cDNA (1256–2303 in Fig. 1*A*) was used as a probe for Southern analysis. According to the sequence of the human IRS-1 cDNA and gene, there are no *Bam*HI, *Eco*RI, or *Hind*III sites and only one *Kpn*I site (nt 1341) in this region; therefore, the expected number of bands was one in *Bam*HI, *Eco*RI, and *Hind*III digests, and two in *Kpn*I digests of genomic DNA. Using high stringency conditions, this was exactly the pattern observed (Fig. 5, high stringency). The relatively weak signal produced by the 1.8-kb *Kpn*I band was expected because of the fact that only 85 bp of the ~1-kb probe could hybridize with this fragment. Additional bands were observed in the *Bam*HI and *Kpn*I digests under moderate stringency conditions (Fig. 5, low stringency), which suggests that one or more additional genes may have some homology to IRS-1 in the genome. This will require further study.

Chromosomal localization of human IRS-1 gene. In situ hybridization of the cDNA probe to normal human metaphase spreads revealed one site that was labeled above background. Of 196 grains in 75 cells scored, 31 (15.8%) were found at the distal end of the long arm of chromosome 2 corresponding to band q36–37 (Fig. 6). No other chromosomal sites were labeled above background.

DISCUSSION

IRS-1/pp185 is an endogenous substrate of the insulin receptor observed in most cells after insulin stimulation (5,6,15–18). We have previously cloned this protein from a rat liver cDNA library and shown it to be a novel signal transduction molecule (7). In this report we have cloned IRS-1 cDNA from human skeletal muscle libraries to determine if this signal transduction protein is conserved across species and tissues, and determined the structure and location of its gene.

Human IRS-1 cDNA has high homology with rat IRS-1 cDNA, with the exception of an increase in length of the 5′-untranslated region. In addition, human IRS-1 cDNA

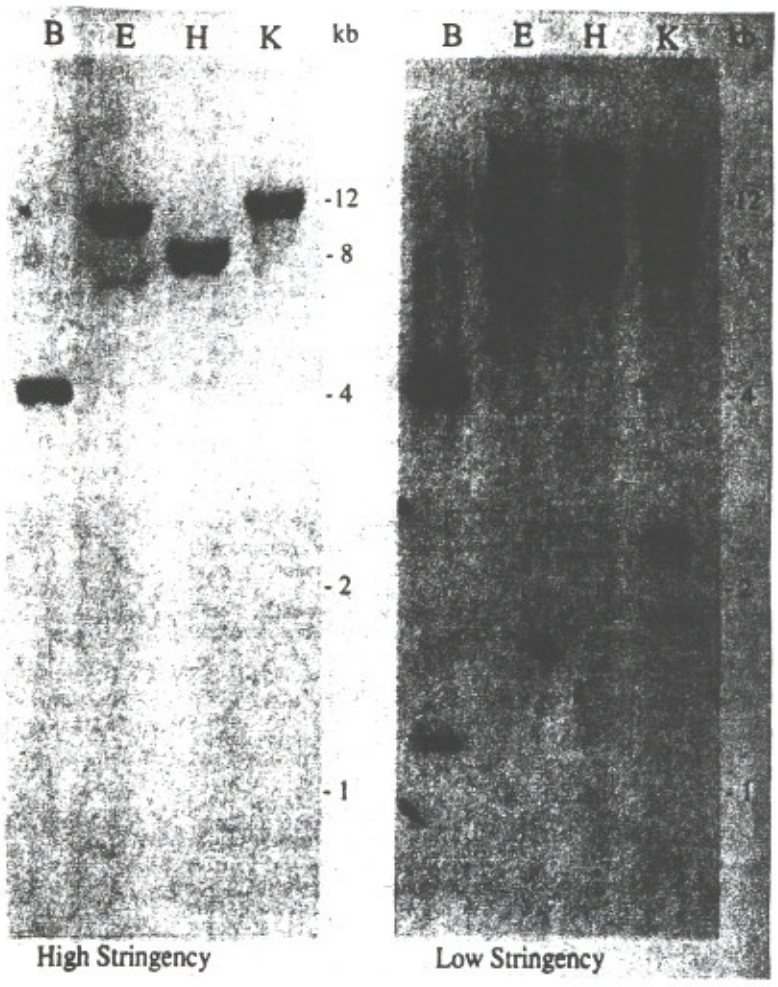


FIG. 5. Southern blot analysis of human IRS-1 gene using high stringency and moderate (low) stringency conditions. Genomic DNA strands (20 μg), prepared from human leukocytes, were digested with restriction enzymes *Bam*HI (*B*), *Eco*RI (*E*), *Hind*III (*H*), and *Kpn*I (*K*), divided into two aliquots of 10 μg each, and Southern blot analysis was performed as described in METHODS.

has a poly(A)$^+$ tail just after the poly(A)$^+$ adenylation signal, which indicates that this signal is functional. In the 5′-untranslated region, there are four potential Sp1 transcription factor–binding sites and a TC-rich region that may regulate IRS-1 gene expression (32–34). There are also TC-rich regions and seven potential Sp1-binding sites in the promoter region of the hIR gene (38), four of which are known to play an important role in its expression (39). Thus, it is possible that the expression of IRS-1 and the insulin receptor is regulated, at least in part, by similar mechanisms.

The deduced amino acid sequence of human skeletal muscle IRS-1 is 88% identical with that of rat liver IRS-1, which indicates that this protein is highly conserved across species and tissues. More interestingly, all putative Tyr phosphorylation sites are completely conserved in these two species, including 6 YMXM and 3 YXXM sites. YMXM and YXXM motifs are known to bind to the SH2 domain of the 85,000-M_r subunit of PtdIns-3 kinase and mediate signal transduction (14,40). Synthetic peptides designed from these YMXM motifs are Tyr phosphorylated by the insulin receptor Tyr kinase in vitro with low K_m (high affinity) (41), and in their phosphotyrosine forms, they bind to the SH2 domain of the PtdIns-3 kinase 85,000-M_r subunit in vitro (M.G. Myers, Jr., unpublished observations). Immunoprecipitation of cell extracts with anti-IRS-1 antibodies coprecipitates an insulin-stimulated PtdIns-3 kinase activity (7). Thus, it appears that IRS-1 is Tyr phosphorylated on its YMXM and/or YXXM motif(s) by the insulin receptor Tyr kinase, binds to the PtdIns-3 kinase 85,000-M_r subunit via SH2 domains, and stimulates this enzyme (42,43). The strong conservation of these motifs between different species and tissues shown herein supports the importance of these regions.

Analysis of the predicted secondary structure (34) of human IRS-1 indicates a very small percentage of α-helix, with the majority of the protein including all of the YMXM and YXXM motifs existing as β-sheet. There are >50 potential Ser/Thr phosphorylation sites in human IRS-1. Several of these exist very close to or even overlapping the YMXM and YXXM motifs. As shown in Fig. 2*B*, two clusters of YMXM and YXXM motifs contain potential Ser/Thr phosphorylation sites arranged in a similar fashion (547–638 and 936-1035). It is possible that this repetitive type of sequence results from some early gene duplication event, and that Tyr phosphorylation of the YMXM and YXXM motifs and Ser/Thr phosphorylation at nearby sites may cross-regulate the activity of one another.

Two putative ATP-binding sites exist in rat IRS-1. In

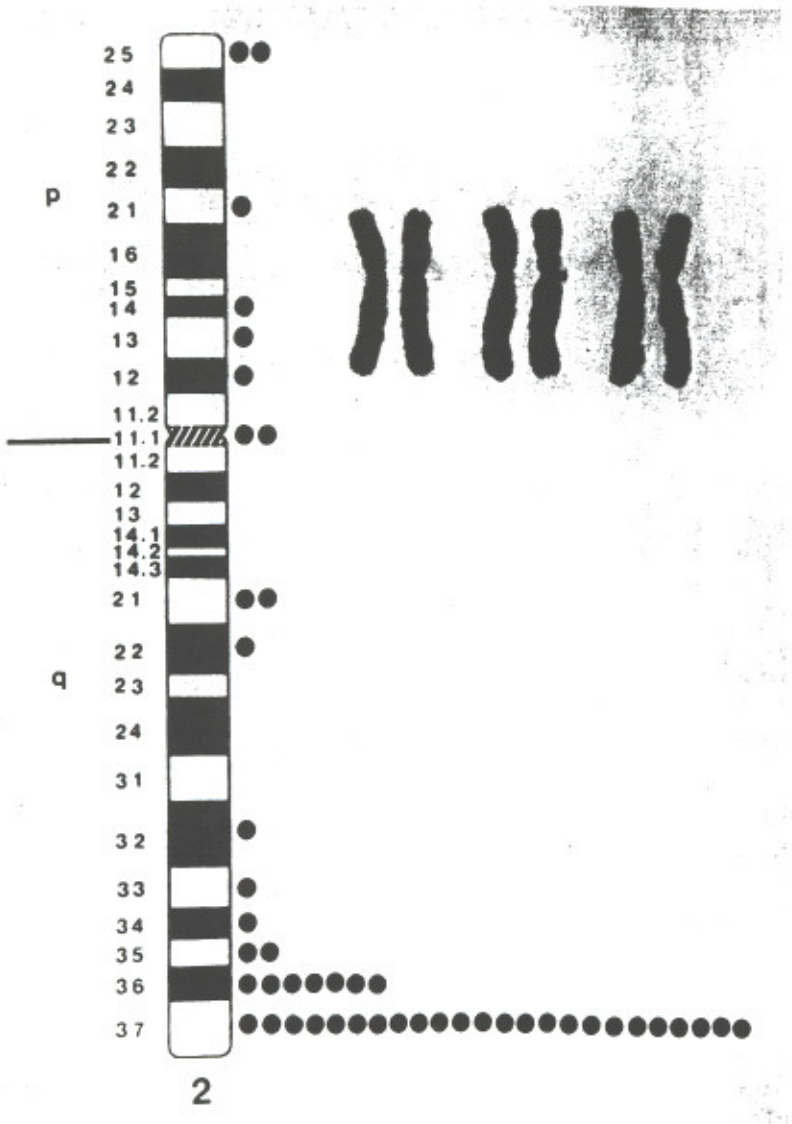


FIG. 6. Chromosomal localizaton of human IRS-1 gene by in situ hybridization. The cDNA probe was nick-translated with [^{3}H]dATP and [^{3}H]dCTP and used for hybridization. Hybridization to human chromosome spreads, posthybridization wash, emulsion autoradiography, and silver grain analysis were conducted as described previously (29). Silver grain distribution along chromosome 2 is shown. Three pairs of chromosome 2 illustrate the specific labeling of the IRS-1 gene.

human IRS-1, only the NH_2-terminal site is conserved, whereas the COOH-terminal site lacks the third Gly in the consensus sequence. Although the function of these nucleotide bindings sites is unknown, in view of the many putative phosphorylation sites in IRS-1 protein, they may have important roles.

The human IRS-1 gene contains the entire 5′-untranslated region, the coding region, and 24 bp of 3′-untranslated region in a single exon. The fact that the whole coding region of IRS-1 exists in a single exon should help the analysis of IRS-1 in studies of insulin resistance using genetic approaches such as single-stranded conformation polymorphisms (44,45) or denaturing gradient gel electrophoresis (46,47).

Southern blot analysis under high stringency conditions was consistent with a single gene for IRS-1; the human IRS-1 gene was localized to the chromosome 2 q36–37 by in-situ hybridization. This corresponds to the central region of mouse chromosome 1 where IRS-1 also has been localized (Y.Z., unpublished observations), and in the human is near the glucagon gene, collagen IV and VI genes, alkaline phosphatase genes, paired-box homeotic gene, and rhabdomyosarcoma-1 genes (48). Southern blot analysis using moderate stringency conditions revealed that at least one additional gene may have some homology with human IRS-1.

The deduced M_r from the amino acid sequence of human IRS-1 is 132,000 M_r, however, human IRS-1 expressed in CHO cells migrates as a 170,000- to 180,000-M_r protein on SDS-PAGE. A similar anomalous migration also has been shown for rat IRS-1 (7), and may be attibutable, in part, to the higher degree of phosphorylation of the molecule. This probably accounts for the varied size estimates of IRS-1 in the literature, which range from 160,000 to 190,000 M_r. After insulin stimulation, human IRS-1 is rapidly Tyr phosphorylated and migrates even higher, i.e., more retarded, on SDS-PAGE gel. This indicates that human IRS-1 is a good substrate of insulin receptor Tyr kinase and that phosphorylation does change the migration on SDS gels.

On Northern blot analysis, two IRS-1 mRNA species of ~6.9 and 6.0 kb are observed using poly(A)$^+$ RNA prepared from every human tissue that was examined. Two mRNA species also are found on Northern blot analysis using poly(A)$^+$ RNA prepared from rat skeletal muscle, but of different size (8.5 and 5.4 kb, E.A., unpublished observations). Using semiquantitative PCR, we find that human IRS-1 is expressed in every fetal tissue examined, which suggests that IRS-1 may play a role in many tissues. Interestingly, both insulin receptor (49) and IRS-1 mRNA (this study) are relatively abundant in fetal brain tissue, raising the possibility that these proteins play some important role in brain development.

Recently an IRS-1–like gene and part of its cDNA was cloned by Nishiyama and Wands (15) using a monoclonal antibody produced against an unknown protein in a human hepatocellular carcinoma. The predicted protein sequence of this cDNA has 99% identity with our human skeletal muscle IRS-1 proteins. However, several differences between our results and their study are noteworthy. First, the 10 amino acid differences: an additional Gly after Gly^{134}, Ser^{362} to Arg, Pro^{384} to Arg, Cys^{436} to Ser, Ala^{620} to Pro, Glu^{729} to Asp, Gly^{822} to Arg, Gly^{1126} to Ala, Asp^{1187} to His, and Cys^{1195} to Ser. Second, the sequence of the 3′-untranslated region of the cDNAs is completely different. The sequence of the 3′-untranslated region reported herein for human IRS-1 has 81% homology with that of rat IRS-1, whereas their clone showed only a 42% homology in this region. Interestingly, the difference between the two cDNAs starts at the exon-intron junction (Fig. 1*A*, base 4771). Thus, it is possible that this difference was generated from differential splicing or a gene rearrangement in the hepatocellular carcinoma cells. A third difference relates to the size of mRNAs detected in Northern blot analysis (our Figs. 3*A* and *B* vs. their Fig. 4 [15]). We have shown two species of mRNAs (6.0 and 6.9 kb) with almost equal abundance in human skeletal muscle, heart, and brain. Nishiyama and Wands (15) found a single 5-kb mRNA in both the hepatocellular carcinoma cells and normal liver. The reason for this discrepancy is also unknown, but may prove an interesting point for further study.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grants DK33201 (C.R.K.) and DK43808 (M.F.W.), the Joslin Diabetes and Endocrinology Research Center Grant DK36836, and the Marilyn M. Simpson Foundation. E.A. is the Mary K. Iacocca Research Fellow.

The authors would like to thank Terri-Lyn Bellman for excellent secretarial assistance; Drs. M. Saad, C. Reynet, and Stacy Kahn for technical assistance; Dr. B.J. Goldstein for helpful advice; Dr. L. Kunkel for the human fetal muscle cDNA library; Dr. S.H. Orkin for the human genomic library; Dr. M. Birnbaum for the pCMV-His plasmid; and Dr. C. Morton for human fetal RNA.

REFERENCES

1. Kahn CR, White MF: The insulin receptor and the molecular mechanism of insulin action. *J Clin Invest* 82:1151–56, 1988
2. Rosen OM: After insulin binds. *Science* 237:1452–58, 1987
3. Ebina Y, Ellis L, Jarmagin K, Edery M, Graf L, Clauser E, Qu HH, Masiarz F, Kan YW, Goldfine ID, Roth RA, Rutter WJ: The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signaling. *Cell* 40:747–58, 1985
4. Ullrich A, Bell JR, Chen EY, Herrera R, Petruzelli LM, Dull TJ, Gray A, Coussens L, Liao YC, Tsubokawa M, Mason A, Seeburg PH, Grunfeld C, Rosen OM, Ramachandran J: Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. *Nature (Lond)* 313:756–61, 1985
5. White MF, Maron R, Kahn CR: Insulin rapidly stimulates tyrosine phosphorylation of a 185,000 *M*, protein in intact cells. *Nature (Lond)* 318:183–86, 1985
6. White MF, Stegmann EW, Dull TJ, Ullrich A, Kahn CR: Characterization of an endogenous substrate of the insulin receptor in cultured cells. *J Biol Chem* 262:9769–77, 1987
7. Sun XJ, Rothenberg PL, Kahn CR, Backer JM, Araki E, Wilden P, Cahill DA, Goldstin BJ, White MF: Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature* (Lond) 352:73–77, 1991
8. Ebina Y, Araki E, Taira M, Shimada F, Mori M, Craik CS, Siddle K, Pierce SB, Roth RA, Rutter WJ: Replacement of lysine residue 1030 in the putative ATP-binding region of the insulin receptor abolishes insulin- and antibody-stimulated glucose uptake and receptor kinase activity. *Proc Natl Acad Sci USA* 84:704–708, 1987
9. Chou CK, Dull TJ, Russel DS, Gherzi R, Lewohl D, Ullrich A, Rosen OM: Human insulin receptors mutated at the ATP-binding site lack protein tyrosine kinase activity and fail to mediate postreceptor effects of insulin. *J Biol Chem* 262:1842–47, 1987
10. Ellis L, Clauser E, Morgan DO, Edery M, Roth RA, Rutter WJ: Replacement of insulin receptor tyrosine residues 1162 and 1163 compromises insulin-stimulated kinase activity and uptake of 2-deoxyglucose. *Cell* 45:721–32, 1986
11. Maegawa H, Olefsky JM, Thies S, Boyd D, Ullrich A, McClain DA: Insulin receptor with defective tyrosine kinase inhibit normal receptor function at the level of substrate phosphorylation. *J Biol Chem* 263:12629–37, 1988
12. White MF, Livingston JN, Backer JM, Lauris V, Dull TJ, Ullrich A, Kahn CR: Mutation of the insulin receptor at tyrosine 960 inhibits signal transmission but does not affect its tyrosine kinase activity. *Cell* 54:641–49, 1988
13. Rothenberg PL, Lane WS, Karasik A, Backer JM, White MF, Kahn CR: Purification and partial sequence analysis of pp185, the major cellular substrate of the insulin receptor tyrosine kinase. *J Biol Chem* 266:8302–11, 1991
14. Koch CA, Anderson D, Moran M, Ellis C, Pawson T: SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* 252:668–73, 1991
15. Nishiyama M, Wands JR: Cloning and increased expression of an insulin receptor substrate-1–like gene in human hepatocellular carcinoma. *Biochem Biophys Res Commun* 183:280–85, 1992
16. Kadowaki T, Koyasu S, Nishida E, Tobe K, Izumi T, Takaku F, Sakai H, Yahara I, Kasuga M: Tyrosine phosphorylation of common and specific sets of cellular proteins rapidly induced by insulin, insulin-like growth factor I, and epidermal growth factor in an intact cell. *J Biol Chem* 262:7342–50, 1987
17. Shemer J, Adamo M, Wilson GL, Heffez D, Zick Y, LeRoith D: Insulin and insulin-like growth factor-1 stimulate a common endogenous phosphoprotein substrate (pp185) in intact neuroblastoma cells. *J Biol Chem* 262:15476–82, 1987
18. Momomura K, Tobe K, Seyama Y, Takaku F, Akanuma Y, Kasuga M: Insulin-induced tyrosine phosphorylation in intact rat adipocytes. *Biochem Biophys Res Commun* 155:1181–86, 1987
19. Meyers Jr. MG, Backer JM, Siddle K, White MF: The insulin receptor functions normally in Chinese hamster ovary cells after truncation of the C-terminus. *J Biol Chem* 266:10616–23, 1991
20. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–67, 1977
21. Davis LG, Dibner MD, Battey JF: pBR 322 colony hybridization. In *Basic Methods in Molecular Biology*. Davis LG, Dibner MD, Battey JF, Eds. New York, Elsevier, 1986, p. 227–29
22. Hartman SC, Mulligan R: Two dominant-acting selective markers for gene transfer studies in mammalian cells. *Proc Natl Acad Sci USA* 85:8047–51, 1988
23. Sambrook J, Fritsch EF, Maniatis T: Plasmid vectors. In *Molecular Cloning*. Ford N, Nolan C, Ferguson M, Eds. 2nd ed. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, 1989, p. 1.12–1.110
24. Southern PJ, Berg P: Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J Mol Appl Genet* 1:327–41, 1982
25. Ebina Y, Edery M, Ellis L, Standring D, Beaudoin J, Roth RA, Rutter WJ: Expression of a functional human insulin receptor from a cloned cDNA in Chinese hamster ovary cells. *Proc Natl Acad Sci USA*

82:8014–18, 1985
26. Towbin H, Staehlin J, Gordon J: Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci USA* 76:4350–54, 1979
27. Gussow D, Rein R, Ginjaar I, Hochstenbach F, Seemann G, Kottman A, Ploegh HL: The human β_2 microglobulin gene: primary structure and definition of the transcription unit. *J Immunol* 139:3132–38, 1987
28. Wang AM, Doyle MV, Mark DF: Quantitation of mRNA by the polymerase chain reaction. *Proc Natl Acad Sci USA* 86:9717–21, 1989
29. Yang-Feng TL, Floyd-Smith G, Nemer M, Drouin J, Francke U: The pronatriodilatin gene is located on the distal short arm of human chromosome 1 and on mouse chromosome 4. *Am J Hum Genet* 37:1117–28, 1985
30. Kozak M: Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44:283–92, 1986
31. Kozak M: The scanning model for translation: an update. *J Cell Biol* 108:229–41, 1989
32. Camilloni G, Seta FD, Negri R, Mauro ED: Structure of RNA polymerase II promoters. *J Biol Chem* 261:6145–48, 1986
33. Kadonaga JT, Carner KR, Masiarz FR, Tjian R: Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA-binding domain. *Cell* 51:1079–90, 1987
34. Courey AJ, Holtzman DA, Jackson SP, Tijian R: Synergistic activation by the glutimine-rich domains of human transcription factor Sp1. *Cell* 59L827–36, 1989
35. Garnier J, Osguthorpe DJ, Robson B: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 120:97–120, 1978
36. Kennelly PJ, Krebs EG: Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J Biol Chem* 266:15555–58, 1991
37. Hanks SK, Quinn AM, Hunter T: The protein kinase family: conserved features and deduced phylogeny of the catalytic domain. *Science* 241:42–52, 1988
38. Araki E, Shimada F, Uzawa A, Mori M, Ebina Y: Characterization of the promoter region of the human insulin receptor gene. *J Biol Chem* 262:16186–91, 1987
39. Araki E, Murakami T, Shirotani T, Kanai F, Shinohara Y, Shimada F, Mori M, Schichiri M, Ebina Y: A cluster of four Sp1 binding sites required for efficient expression of the human insulin receptor gene. *J Biol Chem* 266:3944–48, 1991
40. Cantley LC, Auger KR, Carpenter C, Duckworth B, Graziani A, Kappeller R, Soltoff S: Oncogene and signal transduction. *Cell* 64:281–302, 1991
41. Shoelson SE, Chatterjee S, Chaudhuri M, White MF: YMXM motifs of IRS-1 define substrate specificity of the insulin receptor kinase. *Proc Natl Acad Sci USA* 89:2027–31, 1992
42. Backer JM, Myers MG, Shoelson SE, Chin DJ, Sun X-J, Miralpeix M, Hu P, Margolis B, Skolnik Y, Schlessinger J, White MF: Phosphatidylinositol 3′-kinase is activated by association with IRS-1 during insulin stimulation. *EMBO J* 11:3469–79, 1992
43. Lavan BE, Huhne MR, Garner CW, Anderson D, Reedijk M, Pawson T, Leinhard GE: The association of insulin-elicited phosphotyrosine proteins with src homology 2 domains. *J Biol Chem* 267:11631–36, 1992
44. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T: Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* 86: 2766–70, 1989
45. O'Rahilly S, Choi WH, Patel P, Turner RC, Flier JS, Moller DE: Detection of mutations in insulin-receptor gene in NIDDM patients by analysis of single-stranded conformation polymorphisms. *Diabetes* 40:777–82, 1991
46. Fisher S, Lerman L: DNA fragment differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theoly. *Proc Natl Acad Sci USA* 80:1579–83, 1983
47. Gray MR: Detection of DNA sequence polymorphism in human genomic DNA by using denaturing gradient gel blots. *Am J Human Genet* 50:331–46, 1992
48. Genome Data Base (Baltimore, MD), 1992
49. Lowe WL Jr, Boyd FT, Clarke DW, Raizada MK, Hart C, LeRoith D: Development of brain insulin receptor: structural and functional studies of insulin receptors for whole brain and primary cell cultures. *Endocrinology* 119:25–35, 1986