Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 9732–9736, December 1989
Biochemistry

Characterization of the mouse insulin receptor gene promoter

(3T3-L1 cells/preadipocyte/adipocyte/differentiation)

ERIC SIBLEY*, TANIA KASTELIC*, THOMAS J. KELLY†, AND M. DANIEL LANE*‡

Departments of *Biological Chemistry and †Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

Contributed by M. Daniel Lane, September 14, 1989

ABSTRACT **The 5′ flanking region of the mouse insulin proreceptor gene was isolated, and the 5′ boundary of the minimal promoter was mapped. Genomic clones encompassing >30 kilobases of the gene contain the promoter and exons 1 and 2 interrupted by an ≈20-kilobase intron at the codon for amino acid 7 of the α subunit. The nucleotide sequence of a 1.3-kilobase fragment containing 766 base pairs of the 5′ flanking region and the entire first exon was determined. Two major transcription start sites were mapped by S1 nuclease analysis to sites located 469 and 424 nucleotides upstream from the initiation codon for translation. The 5′ terminus of an insulin proreceptor cDNA, isolated from a mouse 3T3-L1 adipocyte cDNA library, corresponds to the 3′-most major start site of transcription. The 5′ deletion mutants of the 5′ flanking region of the proreceptor gene, linked upstream of the bacterial chloramphenicol acetyltransferase reporter gene, were transfected into 3T3-L1 preadipocytes and assayed for promoter activity. The 5′ boundary of the minimal promoter, which directs unexpectedly high levels of reporter gene expression, maps to a region 22 base pairs upstream from the 3′-most major transcription start site.**

The mature insulin receptor is a transmembrane allosteric enzyme composed of two types of subunits (i.e., α and β subunits) (1, 2). The interaction of insulin with its binding site in the α subunit activates a cytoplasmic tyrosine kinase catalytic domain in the β subunit (3–5). Both the α and β subunits of the insulin receptor are derived through proteolytic cleavage of a common proreceptor precursor (2, 4, 5). The amino acid sequences (deduced from nucleotide sequences of the corresponding cDNAs) for the human (4, 5) and mouse (6) proreceptors are highly conserved; they exhibit about 95% identity at the amino acid level. Recently, the structure of the human insulin proreceptor gene was reported (7). The human gene spans >120 kilobases (kb) and has 22 exons.

Most cells have the ability to regulate the number of functional insulin receptors they possess in response to changes in physiological or developmental state (8). For example, when mouse 3T3-L1 preadipocytes differentiate into adipocytes, there is a 10- to 20-fold increase in the number of cellular insulin receptors, due largely to an increased rate of receptor synthesis (8). As an initial step in investigating the transcriptional control of expression of the insulin proreceptor gene during and after the differentiation of 3T3-L1 preadipocytes into adipocytes, we set out to isolate the promoter-containing 5′ end of the mouse proreceptor gene.

The present communication reports the isolation and characterization§ of the 5′ end (>30 kb) of the mouse insulin proreceptor gene, including a large segment of the 5′ flanking region containing the promoter.

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

EXPERIMENTAL PROCEDURES

Cell Culture. Mouse 3T3-L1 preadipocytes were maintained and induced to differentiate as described (8).

Isolation and Analysis of RNA and DNA. Total cellular RNA was isolated (9) from 3T3-L1 cells during the course of differentiation. For Northern blot analysis, RNA was electrophoresed on 1% agarose/2.2 M formaldehyde gels and transferred (9) to Hybond-N membranes (Amersham). Membranes for Northern blots were hybridized with DNA probes (9) at 42°C in 4× SSC (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7)/5× Denhardt's solution (1× Denhardt's = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin)/10 mM sodium phosphate/10 mM sodium pyrophosphate/0.1% SDS/50% (vol/vol) formamide. The blots were washed at high stringency in 0.1% SDS/0.1× SSC at 65°C and then were autoradiographed.

Isolation of 3T3-L1 Adipocyte cDNA Clones. A mouse placental cDNA library (≈10^6 clones in λgt10) (10) was screened by plaque hybridization with a 230-base-pair (bp) *Sma* I fragment from the 5′ end of a 765-bp mouse placental cDNA, λmIR1 (provided by S. Seino and D. Steiner, University of Chicago), which encodes amino acids 40–295 of the mouse insulin receptor α subunit (Fig. 1). A single 1.3-kb cDNA, λmIR2, was isolated from this screening (Fig. 1). To isolate cDNAs that might extend further 5′, another cDNA library [≈10^6 clones in λZAP (Stratagene)] prepared with poly(A)$^+$ mRNA from differentiated 3T3-L1 adipocytes (9) was screened with a 0.5-kb *Pst* I cDNA fragment from the 5′ end of mouse placental cDNA λmIR2. Ten independent cDNAs were isolated, one of which, λIRc1, extended 109 bp beyond the 5′ end of cDNA λmIR2. The cDNA inserts of positive phage clones were excised, subcloned, and sequenced.

Isolation of Mouse (3T3-L1 Cell) Genomic Clones. A mouse BALB/c genomic DNA library (≈10^6 clones in λCharon 28; provided by P. Leder, Harvard Medical School) was screened with a 0.5-kb *Pvu* II fragment from the 5′ end of mouse placental insulin proreceptor cDNA λmIR1 (Fig. 1). Eight overlapping clones were isolated and characterized by restriction and hybridization analysis. One of these clones, λIRg8, contained a 16-kb insert that extended furthest 5′ (Fig. 1). The 0.5-kb *Pvu* II cDNA probe hybridized to a single 2.2-kb exon-containing *Bam*HI fragment present both in mouse DNA and in λIRg8. This 2.2-kb *Bam*HI fragment of genomic clone λIRg8 was subcloned and sequenced through the 5′ exon–intron boundary of exon 2 by using a sequencing primer complementary to the cDNA coding strand. To isolate exon 1, a mouse 3T3-L1 adipocyte genomic DNA library (≈10^6 clones in λ-EMBL3) (9) was screened with a 0.2-kb *Bam*HI–*Eco*RI fragment (probe A; Fig. 1) from the 5′ end of

Abbreviation: CAT, chloramphenicol acetyltransferase.

‡To whom reprint requests should be addressed at: Department of Biological Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M28869).

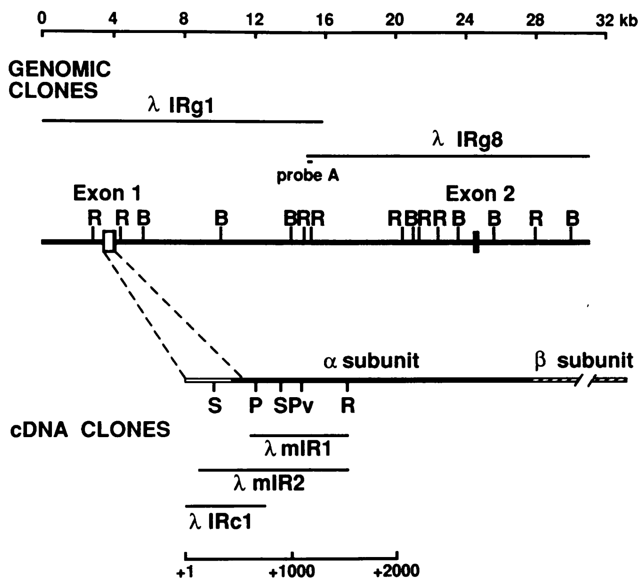

FIG. 1. Map of the 5′ region of the mouse insulin proreceptor gene and cDNA. The positions of cDNAs that encode the 5′ end of the proreceptor mRNA are shown below a schematic representation of the 5′ untranslated (open bar), signal peptide and α-subunit (solid bar), and β-subunit (hatched bar) encoding regions of the proreceptor mRNA. The map of the 5′ region of the mouse insulin proreceptor gene is shown at the top of the figure. The size and location of exon 2 are approximate. Restriction enzyme sites include B, *Bam*HI; R, *Eco*RI; S, *Sma* I; P, *Pst* I; and Pv, *Pvu* II.

genomic clone λIRg8. A single 16-kb genomic clone, λIRg1, was isolated and characterized by restriction and hybridization analysis (Fig. 1). A 0.5-kb *Pst* I probe from the 5′ end of λmIR2 hybridized to a single 1.5-kb exon-containing *Eco*RI fragment present both in mouse DNA and in λIRg1. This 1.5-kb *Eco*RI genomic fragment, which contains exon 1, was subcloned and sequenced.

S1 Nuclease Analysis. A 541-bp genomic DNA restriction fragment (*Hin*dIII–*Stu* I) corresponding to nucleotides −305 to +236 (see Fig. 4) was end-labeled with [γ-^{32}P]ATP and polynucleotide kinase on the noncoding strand at the *Stu* I site (nucleotide +236). S1 nuclease analysis was performed by a modification of the method of Berk and Sharp (11). The probe was added to 15 μg of poly(A)$^+$ RNA from 3T3-L1 adipocytes (9) in 30 μl of 80% formamide/40 mM Pipes, pH 6.4/1 mM EDTA/0.4 M NaCl and incubated at 90°C for 5 min and then at 60°C for 16 hr. S1 nuclease digestion was performed by adding 100 μl of an ice-chilled mixture containing 250 mM NaCl, 2 mM $ZnCl_2$, 30 mM sodium acetate (pH 4.5), and 100 units of S1 nuclease and incubating at 40°C for 40 min. The reaction was stopped by the addition of 80 μl of 4 M ammonium acetate/20 mM EDTA followed by precipitation of the labeled fragments with ethanol. The S1 nuclease digestion products were then analyzed on a 6% acrylamide/7 M urea sequencing gel.

Construction and Transfection of Chimeric Insulin Proreceptor Promoter–Chloramphenicol Acetyltransferase (CAT) Constructs. The 1.0-kb *Eco*RI–*Sma* I genomic DNA fragment corresponding to nucleotides −766 to +252 of the mouse insulin proreceptor gene was modified to generate a fragment bounded by *Bam*HI sites. The 1.0-kb genomic fragment was ligated into a unique *Bam*HI site of the promoterless plasmid pBLCAT3 (13) to generate pIRCAT+ (positive orientation) and pIRCAT− (reverse orientation). The 5′ exonuclease III deletion plasmids were generated by digesting from the vector *Sph* I site of pIRCAT+ (Promega Erase-a-Base system protocol). Individual deletion mutants were sequenced to precisely identify their 5′ termini. 3T3-L1 preadipocytes were transfected with the chimeric promoter–CAT construct DNAs (30 μg per 10-cm dish of 20% confluent cells) and were assayed for CAT activity (expressed per mg of cell extract protein) after 2 days (12).

RESULTS

Isolation of cDNAs Corresponding to the 5′ Untranslated Region of Mouse 3T3-L1 Adipocyte Insulin Proreceptor mRNA. To identify the 5′ end of the mouse insulin proreceptor gene, we sought to isolate cDNA(s) that encode the entire 5′ untranslated region of the receptor message. Initially, a mouse placental cDNA library was screened with a mouse cDNA probe corresponding to amino acids 40–116 of the receptor's α subunit. A single 1.3-kb cDNA, λmIR2, was isolated and sequenced. This cDNA extends 315 bp upstream of an in-frame ATG initiation codon and encodes the first 294 amino acids of the receptor's α subunit (Fig. 1). To isolate cDNAs that might extend farther 5′, a cDNA library prepared with mRNA from differentiated 3T3-L1 adipocytes (9) was screened with a probe from the 5′ end of λmIR2. One of the cDNAs isolated, λIRc1, extends an additional 109 bp beyond the 5′ end of cDNA λmIR2 (Fig. 1) and corresponds exactly to the genomic sequence (see below).

The 3′ end of λIRc1 overlaps cDNA clones described previously (6) that span the entire coding region of the mouse insulin proreceptor. By using one of these cDNAs (i.e., λIRc11, which corresponds to the C-terminal region of the receptor and includes part of the 3′ untranslated region of the message), it was shown that 3T3-L1 cells express two major species of insulin proreceptor mRNA, one of 7.5 kb and another of 9.5 kb (Fig. 2). The proreceptor mRNAs are slightly larger than those reported by Ullrich *et al.* (4). The combined sequence of the entire 5′ untranslated region (424 nucleotides; Fig. 3) and that of the proreceptor coding region (4115 nucleotides; ref. 6) comprise only 4.5 kb. Therefore, based on the sizes of the proreceptor mRNAs, it appears that both mRNA species possess large 3′ untranslated regions. Consistent with our earlier investigations (8) showing that the cellular level and rate of synthesis of the insulin receptor rise dramatically during differentiation of 3T3-L1 preadipocytes, the levels of both proreceptor messages increase markedly

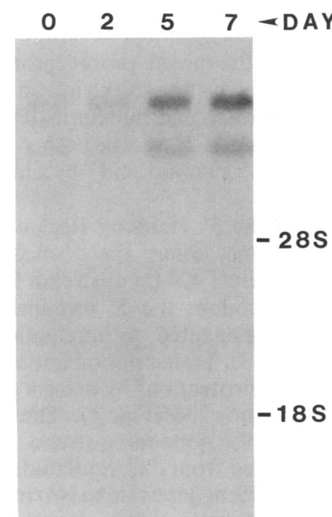

FIG. 2. Northern blot analysis of mouse insulin proreceptor mRNAs expressed during differentiation of 3T3-L1 preadipocytes. Total RNA (50 μg) isolated from 3T3-L1 cells at different times (days 0, 2, 5, and 7) during the course of differentiation was subjected to electrophoresis in an agarose/formaldehyde gel and then transferred to Hybond-N. The RNA blots were hybridized with the cDNA probe λIRc11 (6).

```
 -766 CCACGGGAGTCAAAAAAATAGTAAAAAGGAGTAAAGAAAGACAAAATGTGCTTGGTTGAATCT
 -703 GGCATCTAGGGCATATTGGGGTGTTTATGTCCTGGGGTATCTGGCGGCATTTCCAGACTTTAAAATTGTT
 -633 ATTTCGGGGTAATTTGGGGTGTCTCGCTCTGGCGGAGATCTAGAGTGTGCTCCAGACCTTTAAATTGGAA
 -563 CCTCGGTGTTACTTGTGTTTGCTGCAGTCAGGAGGGGACCTGGGGGCGTCCCTAGGCTTTTAAACTGGAA
 -493 CCTCAGGCCTGTTTAGTGGTTTTCGAGTCAGGCGGGGATCTGGGGACCCCTAGGTTTTAAAATCGGAACT
 -423 TTGGGACTCTTTGGGGGCGTTCGTGTCAGGTGCGGATATGGGCAGCTCCGGACGTTTGACAGGTGAGTCT
 -353 TAGAGTTGTTTGGGTGTTTATGTTACACTGGTATCTGAGAACGCTTACAAGCTTTTAAATCGTAATCTTT
 -283 CTAAGCCAGCAGTTCGGGGTGTTGGGATTGGAAGAATCTGAAGTGAGCCTCGGGTCTCTGAAGTTGCGGG
 -213 TTTTACGGGGCAAGGGGGAAAGGGGCCGAGGAGGCGAGGAGGCGAGGAGGCGTGGCGTTGCCAGCTCTGG
 -143 CCACGCCTCCCCGAGCTCTGCTGGCTTTCTCCCCCTCCCCCCCCAGCGGGCCTCACTGGAGCGCACAGTC
  -73 CCTTCCCTGTCTTAATCCGCGCCACCTTTTCCCTTGGCCCGCGCGGGCCCAGCTGACGGGCTGCGTTGTT
   -3 TACGGGCTGGAGTCGAGCCCTAGCTCCCGCCGCCACCACCAGCCCAGGTGCCCTCCAGTGCGCGCGTCCC
  +68 TCTGTCTCAGACTCCACACAACGCGACCCACCCAGAGCTGCACGGAGGGCAGAAGCAGGAGACCCGGACA
 +138 GGAGACGCACCGCGGTGAGCTCTGGACTCTGGGATTGCGGGCACGGACCGGGCCTGGGTGACCTGCGGG
 +208 CCGCGCCACGGTGTTGCTTGCGGCCGAGGCCTCTGTGCTCTTCCCGGGACTGTCCCCAGGGCCCTCTAGG
 +278 CTGGAGAGCTGCGGCCTGTGAGCCACGGGCGTGGAAGAGAAGGACGTGCGGCCCCGAGCGCCTCTCCAGA
 +348 GACCTTCTCACGGAGTATGTCCCCAGTAGGCCGGCGTGGCGTGCTCTGATCGCCGGGGTCCCAGCACTCC

                -27                         -20
                Met Gly Phe Gly Arg Gly Cys Glu Thr Thr Ala Val Pro Leu Leu
 +418   TACTGCT ATG GGC TTC GGG AGA GGA TGT GAG ACG ACG GCT GTG CCA TTG CTG

                -10                                     +1
        Val Ala Val Ala Ala Leu Leu Val Gly Thr Ala Gly His Leu Tyr Pro Gly
 +470   GTG GCC GTG GCC GCG TTG CTG GTG GGC ACA GCC GGC CAC CTG TAC CCT GGA

         +6
        Glu
 +521   GAG GGTAAGTCTGG
                Intron
```

FIG. 3. Nucleotide sequence of the 5′ flanking region and exon 1 of the mouse insulin proreceptor. The 5′ end of the cDNA is designated as nucleotide +1 (boldfaced and double underlined) of the gene. The two major transcription start sites mapped by S1 nuclease analysis (see Fig. 4) are indicated by asterisks (★). Several minor start sites are denoted by solid dots (·). Two pyrimidine-rich regions and a purine-rich region are indicated by solid underlines and overlines, respectively. The dashed overline indicates three tandem repeats of a CGAGGAGG sequence.

when 3T3-L1 preadipocytes undergo differentiation into adipocytes (Fig. 2).

Structure of the 5′ Region of the Mouse Insulin Proreceptor Gene. Two overlapping genomic clones (λIRg1 and λIRg8), which span ≈30 kb, were isolated from two mouse genomic libraries (see *Experimental Procedures*). The first and second exons were mapped by using cDNA fragments as hybridization probes. The sizes and relative locations of the exon-containing fragments corresponded precisely to those predicted by Southern analysis. Exons 1 and 2 are interrupted at amino acid 7 by a consensus 5′ splice donor site (GAG Ggtaagt) and a consensus 3′ splice acceptor site (ttctagTG TGC). The 5′ region of the mouse proreceptor gene consists of exons 1 and 2 separated by an ≈20-kb intron (Fig. 1). Thus, the 5′ regions of the mouse and human insulin proreceptor genes are similarly organized, and in both cases exon 1 is interrupted at the codon for amino acid 7 by a large first intron (7).

Characterization of the 5′ Flanking Region of the Mouse Insulin Proreceptor Gene. Since the 5′-most proreceptor cDNA (i.e., λIRc1) extends 424 bp upstream from the translation initiation ATG codon, the 5′ terminal base of this cDNA is arbitrarily designated as nucleotide +1 in the genomic sequence (Fig. 3). Transcription initiation sites were mapped by S1 nuclease protection¶ by using a genomic probe extending from −305 to +236 (Fig. 4). The probe was 5′ end-labeled with ^{32}P on the noncoding strand and hybridized to poly(A)$^+$ RNA isolated from differentiated 3T3-L1 adipocytes. S1 nuclease treatment gave rise to two major protected fragments of 281 and 234 nucleotides in length and to several minor protected fragments clustered in the same regions. The two major protected fragments correspond to initiation sites located at nucleotides −45 and +3, the latter site being nearly equivalent to the 5′ end of the message encoded by λIRc1 cDNA.

Analysis of the 5′ flanking sequence of the proreceptor gene (Fig. 3) reveals an A+T-rich region at −63, which is located 18 bp upstream of the first major transcription start site. A CCAAT homology is not present. Three tandem repeats of the sequence CGAGGAGG are located between positions −187 and −154. A purine-rich region in the coding strand, which spans the three tandem repeats, is followed downstream by two pyrimidine-rich regions (see Fig. 3). Two direct repeats of the sequence ACGGGCTG are located between nucleotides −18 and −11 and nucleotides −2 and +6. The 5′ region of the mouse insulin proreceptor gene is extremely G+C-rich; the first 300 bp of the 5′ flanking region and the 424-bp 5′ untranslated region have an overall G+C content of 67%. Other known cis-acting regulatory elements were not found.

Localization of Promoter Activity Within the 5′ Flanking Region of the Insulin Proreceptor Gene. To assess promoter activity, the 1.0-kb *Eco*RI–*Sma* I genomic fragment of λIRg1, which includes nucleotides −766 to +252, was fused to the CAT reporter gene in the promoterless expression vector, pBLCAT3 (13), to generate hybrid constructs pIRCAT+ and pIRCAT−, in which the promoter is in the reverse orientation relative to the CAT gene (Fig. 5). The 1.0-kb genomic fragment spans the transcription start sites mapped above by S1 nuclease protection. The chimeric genes were transfected into 3T3-L1 preadipocytes and assayed for CAT activity. Cell extracts from preadipocytes transfected with pIRCAT+ exhibited extremely high levels of CAT activity‖ (i.e., >500-fold higher than the level of CAT activity for cells transfected with pIRCAT− or pBLCAT3, the vector alone).

¶Primer extension analysis generated numerous truncated products apparently due to frequent stops by reverse transcriptase in the G+C-rich 5′ untranslated region of the insulin proreceptor message.

‖CAT activity expressed by pIRCAT+ transiently transfected into 3T3-L1 cells is greater than that expressed by RSV.CAT; the Rous sarcoma virus promoter is a strong promoter in 3T3-L1 cells (12).

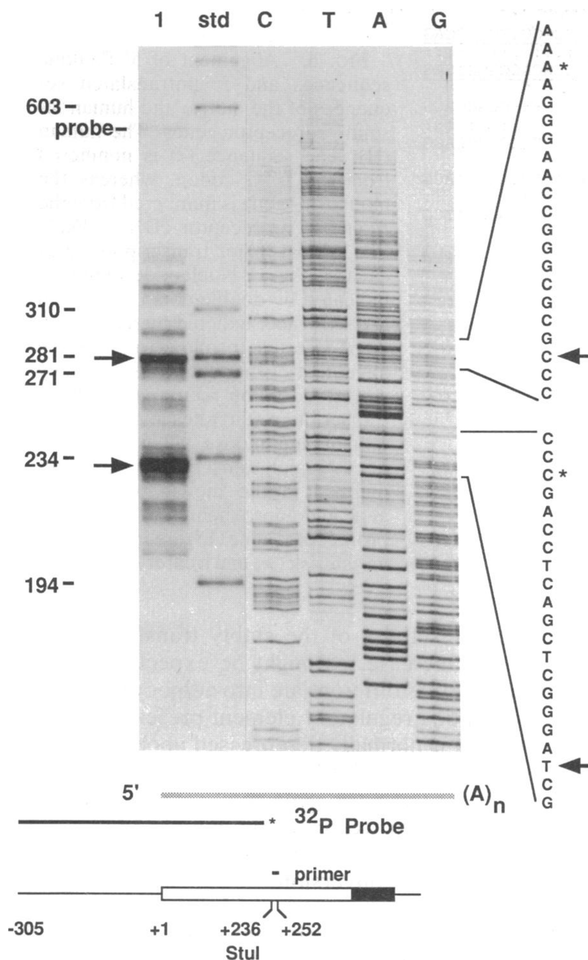

FIG. 4. Localization of major transcription start site(s) within the mouse insulin proreceptor gene by S1 nuclease protection analysis. Poly(A)$^+$ RNA (15 μg) from differentiated 3T3-L1 adipocytes was hybridized to a 541-bp *Hin*dIII–*Stu* I genomic fragment 5' end-labeled at the *Stu* I site (+236) as shown in the diagram. The RNA·DNA hybrids were digested by S1 nuclease and electrophoresed through a sequencing gel. Sizes of the protected bands were determined from the adjacent genomic sequence, which was generated by using an oligonucleotide primer with a 5' terminus located at nucleotide +252. The major transcription start sites (indicated by arrows) were mapped 16 nucleotides upstream to the sequence positions indicated by an asterisk (★). The single-stranded genomic DNA migrates slightly faster than the *Hae* III-digested ϕX174 DNA markers, presumably due to strong secondary structure. The start site locations and sequence were confirmed on subsequent gels by using a primer at nucleotide +236 (data not shown).

Promoter activity was mapped by comparing CAT expression driven by a series of 5' promoter deletion mutants with CAT expression driven by the full-length parental pIRCAT+ construct (Fig. 5). The 5' exonuclease III deletion mutants of the proreceptor promoter were generated from pIRCAT+, and the 5' termini of the deleted inserts were mapped (by sequencing) to nucleotide positions −492, −305, −62, −22, +13, and +34. The parental pIRCAT+ construct and deletion mutants pIRCAT/−492, -/−305, -/−62, and -/−22, which have 5' boundaries upstream of the 3'-most major transcription start site, all promote high levels of CAT activity (i.e., 227–574 pmol/mg per hr) (Fig. 5). In contrast, 5' deletions that extend beyond the 3'-most major transcription start site (i.e., pIRCAT/+13 and -/+34) promote extremely low levels of CAT activity (i.e., 3–9 pmol/mg per hr) (Fig. 5). Thus, the 5' boundary for the minimal promoter lies between nucleotides −22 and +13. These results are also consistent with mapping of the 3'-most major transcription

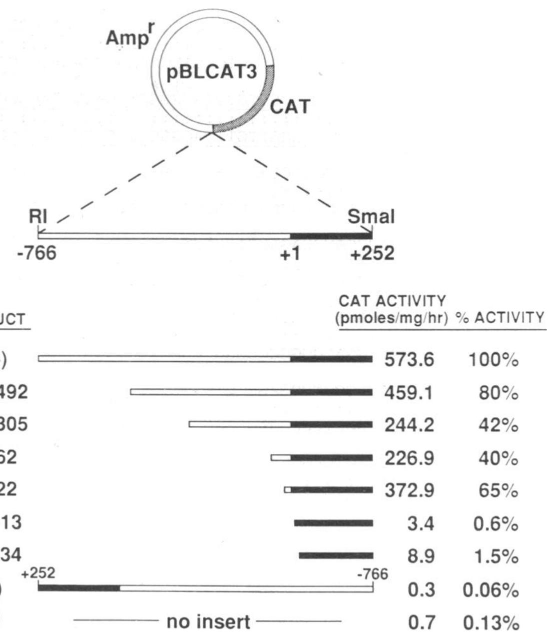

FIG. 5. Promoter activity in 3T3-L1 cells of various 5' flanking segments of the mouse insulin proreceptor gene linked to the CAT reporter gene. A 1.0-kb *Eco*RI–*Sma* I fragment of the proreceptor gene was subcloned into the promoterless plasmid pBLCAT3 to generate pIRCAT+ and pIRCAT− (with the promoter segment in the reverse orientation). The 5' deletion constructs were generated by exonuclease III digestion from the 5' end of pIRCAT+. 3T3-L1 preadipocytes were transiently transfected with the CAT constructs and assayed for CAT activity after 48 hr. The 5' flanking sequence upstream of nucleotide +1 is represented by an open box. The 5' untranslated region downstream of nucleotide +1 is represented by a solid box. Amp^r, ampicillin-resistance gene.

start site (see Figs. 3 and 4) since promoter activity is virtually abolished when 5' deletions extend beyond this site (Fig. 5).

DISCUSSION

The amino acid sequences of the mouse and human insulin proreceptors are nearly identical; they exhibit 95% overall amino acid sequence identity (4–6). Likewise, the structures of the 5' ends of the mouse and human (7, 14, 15) insulin proreceptor genes are quite similar. In both genes, a splice junction interrupts the codon for amino acid 7 in the mature α subunit separating exons 1 and 2 with a large first intron [i.e., ≈20 kb in the mouse proreceptor (Fig. 1) and at least 20 kb in the human proreceptor] (7). As illustrated in Fig. 6, there is also substantial sequence homology within the 5' untranslated and 5' flanking regions of both genes, which are unusually G+C-rich. Alignment of the sequence of the 5' untranslated region of the mouse gene (nucleotides +1 to +424) with respect to the comparable region of the human gene upstream from the ATG codon reveals an overall sequence identity of 62% (Fig. 6). An even higher degree of sequence identity (i.e., 86%) between the two genes occurs within the 90-bp region surrounding the major transcription start site of the mouse proreceptor gene (see underlined sequences in Fig. 6). This region includes the minimal promoter sequence required for near-maximal expression of the proreceptor promoter–CAT construct transfected into 3T3-L1 cells (see Fig. 5 and below).

There is some uncertainty regarding the location of transcription start sites in the human insulin proreceptor gene. Transcription start sites in the human gene have been mapped by different investigators to nucleotides located 230 bp (14); 276, 282, and 283 bp (7); as well as 250, 260, 420, 430, 440, 480,

```
M   -91 TCACTGGAGCGCACAGTCCCTTCCCTGTCTTAATCCGCGCCACCTTTTCCCTTGGCCCGCGCGGGCCCAG
H  -550 TCTCCCGGGCGCAGAGTCCCTTCCTAGGCCAGATCCGCGCCGCCTTTTCCCGCAGCCCGCACGGGCCCAG
                                 +1
M   -21 CTGACGGGCTGCGTTGTTTACGGGCTGGAGTCGAGCCCTAGCTCCCGCCGCCA---------CCACCAGC
H  -480 CTGACGGGCCGCGTTGTTTACGGGCCCGAGC--AGCCCTCTCTCCCGCCGCCCGCCCGCCACCCGCCAGC

M   +41 CCAGGTGCCCTCCAG-------------------TGCGCGCGTCCCTCTGTCTCAGACTCCACACAACG
H  -412 CCAGGTGCCCGCCCGCCAGTCAGCTAGTCCGTCGGTCCGCGCGTCCCTCTGTCCCGGAGCCCGCAGATCG

M   +91 CGACCCACCCAGAGCTGCACGGAGGGCAGAAGCAGGAGACCCGGACAGGAGACGCACCGCGGT--GAGCT
H  -342 CGACCCAGAGCGCGCGGGGCCGAGAGCCGAGAGACAGTCCCGGGCGCAGCGGAGCTCCGGGCCCCGAGAT

M  +159 CTGGACTCTGGGATTGCGG-GCACGGGACCGGGCCTGGGTG--ACCTGCG----GGCCGCGCCACGGTGT
H  -272 CCTGGGACGGGGCCCGGGCCGCAGCGGCCGGGGGGTCGGGGCCACCACCGCAAGGGCCTCCGCTCAGTAT

M  +222 TGCTTGCGGCCGAGGCCTCT-GTGCTCTTCCCGGGACTGTCCCCAGGGCCCTCTA-GGCTGGAGAGCTGC
H  -202 TTGTAGCTGGCGAAGCCGCGCGCGCCCTTCCCGGGGCTG-CCTCTGGGCCCTCCCCGGCAGGGGGGCTGC

M  +290 GGCCTGTGAGCCACGGGCGTGGAAGAGAAGGACGTGCGGCCCCG-AGCGCCTCTCCAGAGACCTTCTCAC
H  -133 GGCC-GCGGGTCGCGGGCGTGGAAGAGAAGGACGCGCGGCCCCCCAGCGCCTCTTGGGTGGCCGCCTC--

M  +359 GGAGTATGTCCCCAGTAGGCCGGCGTGGCGTGCTCTGATCGCCGGGGTCCCAGCACTCCTACTGCTATGG
H   -66 GGAGCATGACCCCCGCGGGCCAGCGCCGCGCGCTCTGATCCGAGGAGACCCCGCGCTCCCGCAGCCATGG
                                                                        +1
```

FIG. 6. Alignment of 5′ flanking sequences and 5′ untranslated sequences of the mouse and human insulin proreceptor genes. The human (H) gene sequence (7) is numbered from the ATG codon, whereas the mouse (M) gene is numbered from the 5′ end of proreceptor cDNA λIRc1, the 3′-most major transcription start site (see Fig. 1). Nucleotide identity is indicated by double dots (:). Gaps in the sequence, indicted by dashes (–), were inserted to allow for maximal alignment of the two genes. A highly conserved 90-bp region surrounding transcription start sites mapped in the human and mouse genes is underlined. Major transcription start sites mapped in the mouse proreceptor gene and approximated in the human proreceptor gene (15) are indicated by an asterisk (★) and a caret (^), respectively.

490, 510, and 520 bp (15) upstream of the ATG translation initiation codon. In none of these cases have full-length cDNAs been reported. Based on the location of the 5′ end of a mouse (3T3-L1 cell) cDNA, which extends 424 bp upstream of the ATG codon, and on the results of S1 nuclease analysis, we have identified the major transcription start site for the mouse proreceptor gene (Figs. 3 and 4). A second transcription start site, located 46 bp further upstream, and several minor sites clustered in the same region were also detected (Figs. 3 and 4). The transcription start sites within the mouse gene and four of the upstream start sites in the human gene, identified by Mamula *et al.* (15), map to the highly conserved 90-bp region of the 5′ flanking sequence described above (Fig. 6). The location of transcription initiation sites within this region of the mouse and human genes is consistent with our studies on the localization of promoter activity within the 5′ flanking region of the mouse gene.

A 5′ flanking segment of the mouse proreceptor gene directs high levels of transcription in 3T3-L1 cells when fused upstream of the CAT reporter gene (Fig. 5). This segment of the proreceptor gene drives CAT expression at a level greater than that driven by the strong Rous sarcoma virus viral promoter‖ (V. Yang and M.D.L., unpublished results). By analyzing 5′ deletions of the proreceptor gene promoter, we found that the 5′ boundary for minimal promoter activity is located between nucleotides −22 and +13, which span the major transcription start site (Fig. 5; also see Fig. 3). It is surprising that an extremely high level of CAT expression is driven by the minimal proreceptor promoter, which includes only 22 nucleotides of 5′ flanking sequence (Fig. 5) and lacks any recognizable positive cis-acting regulatory elements. Conceivably, sequences immediately surrounding and including the major transcription start site of the insulin proreceptor gene are sufficient to direct maximal expression. Such an "initiator" transcriptional control element, comprised of 17 bp, mediates transcription initiation of the lymphocyte-specific terminal deoxynucleotidyltransferase gene and contains within itself a transcription start site (16). Alternatively, cis-acting regulatory sequences capable of activating transcription of the proreceptor gene may lie downstream from the transcription start site (i.e., within the first 252 bp of the 5′ untranslated region present in the chimeric genes) (Fig. 5). In this context it is of interest that the promoter of the pIRCAT+ chimeric gene is active when transiently (Fig. 5) or stably (E. Sibley and M.D.L., unpublished results) transfected into 3T3-L1 preadipocytes. Moreover, promoter activity of the stably transfected chimeric gene does not increase, as might be expected (see Fig. 2) when preadipocytes differentiate into adipocytes. Perhaps, a negative cis-acting regulatory element present in the endogenous gene that is normally derepressed upon differentiation is not present in the pIRCAT+ chimeric gene. Further study of the insulin proreceptor gene may reveal 5′ flanking or intron sequences that regulate the activity of the minimal promoter to account for the differential expression of this gene during adipose conversion.

We thank Vincent Yang, Robert Christy, Michael Hogan (Baylor College of Medicine), Susan Buhrow, and Klaus Kaestner for helpful advice and discussions. We are grateful to Natalie Tumminia for assistance in preparing this manuscript. This work was supported by a research grant from the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (DK 38418). E.S. was supported by a Predoctoral Fellowship from the Henry C. Wellcome Fellowship, State of Maryland.

1. Massague, J., Pilch, P. F. & Czech, M. P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7137–7141.
2. Olson, T. S. & Lane, M. D. (1988) *J. Biol. Chem.* **263**, 7342–7351.
3. Rosen, O. M. (1987) *Science* **237**, 1452–1458.
4. Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petruzzelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y. C., Tsubokawa, M., Masson, C., Seeburg, P. H., Grunfeld, C., Rosen, O. M. & Ramachandran, J. (1985) *Nature (London)* **313**, 756–761.
5. Ebina, Y., Ellis, L., Jarnaglin, K., Edery, M., Graf, L., Clauser, E., Ou, J., Masiarz, F., Kan, Y. W., Goldfine, I. D., Roth, R. A. & Rutter, W. J. (1985) *Cell* **40**, 747–758.
6. Flores-Riveros, J. R., Sibley, E., Kastelic, T. & Lane, M. D. (1989) *J. Biol. Chem.*, in press.
7. Seino, S., Seino, M., Nishi, S. & Bell, G. I. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 114–118.
8. Reed, B. C., Ronnett, G. V., Clements, P. R. & Lane, M. D. (1981) *J. Biol. Chem.* **256**, 3917–3925.
9. Ntambi, J. M., Buhrow, S. A., Kaestner, K. H., Christy, R. J., Sibley, E., Kelly, T. J. & Lane, M. D. (1988) *J. Biol. Chem.* **263**, 17291–17300.
10. Stempien, M. M., Fong, N. M., Rall, L. B. & Bell, G. I. (1986) *DNA* **5**, 357–361.
11. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–732.
12. Yang, V. W., Chirsty, R. J., Cook, J. S., Kelly, T. J. & Lane, M. D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3629–3633.
13. Luckow, B. & Schütz, G. (1987) *Nucleic Acids Res.* **15**, 5490.
14. Araki, E., Shimada, F., Uzawa, H., Mori, M. & Ebina, Y. (1987) *J. Biol. Chem.* **262**, 16186–16191.
15. Mamula, P. W., Kwok-Ying, W., Maddux, B. A., McDonald, A. R. & Goldfine, I. D. (1988) *Diabetes* **37**, 1241–1246.
16. Smale, S. T. & Baltimore, D. (1989) *Cell* **57**, 103–113.