Cloning and characterization of the human and rat islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) genes.

C. C. Martin, L. J. Bischof, B. Bergman, L. A. Hornbuckle, C. Hilliker, C. Frigeri, D. Wahl, C. A. Svitek, R. Wong, J. K. Goldman, et al.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 27, Issue of July 6, pp. 25197–25207, 2001

Printed in U.S.A.

Cloning and Characterization of the Human and Rat Islet-specific Glucose-6-phosphatase Catalytic Subunit-related Protein (IGRP) Genes*

Received for publication, February 19, 2001, and in revised form, March 27, 2001
Published, JBC Papers in Press, April 10, 2001, DOI 10.1074/jbc.M101549200

Cyrus C. Martin‡§, Larry J. Bischof‡, Barbara Bergman¶, Lauri A. Hornbuckle‡‖, Carl Hilliker¶, Claudia Frigeri‡, David Wahl¶, Christina A. Svitek‡, Randall Wong¶, Joshua K. Goldman‡, James K. Oeser‡, Frédéric Leprêtre, Philippe Froguel**, Richard M. O'Brien‡ ‡‡, and John C. Hutton¶ ‡‡**

From the ‡Department of Molecular Physiology and Biophysics, Vanderbilt University Medical School, Nashville, Tennessee 37232, ¶Barbara Davis Center for Childhood Diabetes, University of Colorado Health Sciences Center, Denver, Colorado 80262, and ‖Institut de Biologie-CNRS 8090, Institut Pasteur de Lille, Lille Cedex, France

Islet-specific glucose-6-phosphatase (G6Pase) catalytic subunit-related protein (IGRP) is a homolog of the catalytic subunit of G6Pase, the enzyme that catalyzes the terminal step of the gluconeogenic pathway. Its catalytic activity, however, has not been defined. Since *IGRP* gene expression is restricted to islets, this suggests a possible role in the regulation of islet metabolism and, hence, insulin secretion induced by metabolites. We report here a comparative analysis of the human, mouse, and rat *IGRP* genes. These studies aimed to identify conserved sequences that may be critical for IGRP function and that specify its restricted tissue distribution. The single copy human *IGRP* gene has five exons of similar length and coding sequence to the mouse *IGRP* gene and is located on human chromosome 2q28–32 adjacent to the myosin heavy chain 1B gene. In contrast, the rat *IGRP* gene does not appear to encode a protein as a result of a series of deletions and insertions in the coding sequence. Moreover, rat IGRP mRNA, unlike mouse and human IGRP mRNA, is not expressed in islets or islet-derived cell lines, an observation that was traced by fusion gene analysis to a mutation of the TATA box motif in the mouse/human IGRP promoters to TGTA in the rat sequence. The results provide a framework for the further analysis of the molecular basis for the tissue-restricted expression of the *IGRP* gene and the identification of key amino acid sequences that determine its biological activity.

Glucose-6-phosphatase (G6Pase)[1] is located in the endoplasmic reticulum (ER) and catalyzes the terminal step of the gluconeogenic pathway in liver and kidney. The enzyme is thought to be a multi-subunit complex; however, the exact number of subunits, their stoichiometry, and topological relationships are unclear (1, 2). The 36-kDa G6Pase catalytic subunit spans the membrane multiple times and appears, based on studies using microsomes, to have its catalytic site oriented toward the lumen (1, 2). A model has therefore been proposed in which the G6Pase catalytic subunit is postulated to be associated functionally with a 46-kDa glucose 6-phosphate (G6P) transporter (3) and hypothetical transporters for inorganic phosphate and glucose, which serve to deliver cytosolically generated G6P to the active site and shuttle the reaction products back to the cytosol (1, 2). Rapid kinetic data, on the other hand, favor an alternative model that places the catalytic site within the membrane and ascribes both a transport function and catalytic activity to the 36-kDa catalytic subunit (1, 2). Mutations within the G6Pase catalytic subunit cause glycogen storage disease type 1a (4), which is characterized by severe hypoglycemia in the post-absorptive state, hepatomegaly associated with excessive glycogen deposition, growth retardation, and renal failure (4). In glycogen storage disease type 1b mutations in the G6P transporter give rise to a similar phenotype and additional complications possibly related to independent functions of this molecule in other tissues (4).

Hepatic G6Pase activity is increased in poorly controlled human type 1 and 2 diabetics (5, 6) and in experimental rodent diabetic models (7–11). Along with the elevated activity of other gluconeogenic enzymes, it contributes to an increase in hepatic glucose production and the hyperglycemia that characterizes the disease (5, 6, 12, 13). The change in G6Pase activity has

* This work was supported by a grant from Juvenile Diabetes Foundation International (JDFI) and Vanderbilt Diabetes Core Laboratory Grant P60 DK20593 (to R. O'B.), American Diabetes Association Grant 9901-116 and Barbara Davis Center Diabetes and Endocrinology Research Center Grant P30 DK57516 (to J. C. H.), and by a grant from the Nord-Pas de Calais Region (to F. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF283835 (human IGRP gene excluding the promoter), AF283575 (human IGRP promoter), AF321459-AF321463 (human IGRP exons 1–5, respectively), AF323433 (rat IGRP promoter and exon 1), AF323434-AF323436 (rat IGRP exons 2, 3, and 5, respectively), NM021331 (mouse IGRP cDNA), AF118761 (mouse IGRP promoter), and AF 118762-AF118766 (mouse IGRP exons 1–5, respectively).

§ Supported by Vanderbilt Viruses, Nucleic Acids, and Cancer Training Program 5T32 CA09385-17).

‖ Recipient of Vanderbilt Molecular Endocrinology Training Program Award 5 T 32 DK07563-12).

‡‡ To whom correspondence may be addressed: Dept. of Molecular Physiology and Biophysics, 761 MRB II, Vanderbilt University Medical School, Nashville, TN 37232-0615. Tel.: 615-936-1503; Fax: 615-322-7236; E-mail: richard.obrien@mcmail.vanderbilt.edu (R. O'B.) or Tel.: 303-315-8197; Fax: 303-315-4892; E-mail: john.hutton@uchsc.edu (J. C. H.).

[1] The abbreviations used are: IGRP, islet-specific G6Pase catalytic subunit-related protein; G6Pase, glucose-6-phosphatase; ER, endoplasmic reticulum; G6P, glucose 6-phosphate; ORF, open reading frame; CAT, chloramphenicol acetyltransferase; RIN and INS-1, rat insulinoma-derived cell lines; contig, group of overlapping clones; HIT, hamster insulinoma tumor; bp, base pair(s); kbp, kilobase pair(s); PCR, polymerase chain reaction; RT, reverse transcription; nt, nucleotides; MES, 4-morpholineethanesulfonic acid; aa, amino acid(s).

been attributed to changes in expression of the genes encoding both the G6Pase catalytic subunit and the G6P transporter. The former probably reflects the combined stimulatory effect of glucose (10, 14, 15) and the loss of the inhibitory action of insulin (16, 17). Less is known about the factors that regulate expression of the G6P transporter (18). However, experimental overexpression of either the G6Pase catalytic subunit (19) or the G6P transporter (20) in hepatocytes using recombinant adenovirus leads to enhanced rates of G6P hydrolysis as well as changes in glycogen metabolism.

Most studies show that islets also contain a hydrolytic activity that is specific for G6P but that is present at a lower specific activity than liver (2, 21–25). G6Pase activity is elevated in islets isolated from ob/ob mice resulting in increased glucose substrate cycling (26, 27). The question of whether islet G6Pase activity is catalyzed by the same G6Pase catalytic subunit as in liver, however, has proven controversial (2, 21). Thus, the G6P hydrolytic activity in islets displays distinct kinetic behavior and inhibitor profiles compared with that in hepatic extracts (21).

We recently identified a novel cDNA from mouse β cell-derived cell lines that encodes an islet-specific G6Pase catalytic subunit-related protein (IGRP) (21). IGRP is a putative ER membrane protein that is similar in size (38 kDa), topology, and sequence (~50% identity at the amino acid level) to the G6Pase catalytic subunit (21). The function of IGRP, however, is uncertain since its overexpression in fibroblast or endocrine cell lines does not increase rates of G6P hydrolysis in tissue homogenates (21). The mouse *IGRP* gene has a chromosomal locus that is distinct from the G6Pase catalytic subunit but it has a similar exon/intron structure, suggesting that the genes arose from an ancient gene duplication/transposition event (28). To date, *IGRP* gene expression has only been detected in pancreatic endocrine cells (21), a feature that is reflected in the islet-specific activity of the IGRP promoter as assessed using fusion gene constructs. Thus, the IGRP promoter is inactive in human HepG2 hepatoma cells but is ~150-fold more active than the G6Pase catalytic subunit promoter in hamster insulinoma tumor (HIT) cells (28).

In this paper we report a comparative analysis of the structure of the *IGRP* genes from the mouse, human, and rat and an investigation of their expression at the level of tissue mRNA and promoter activity. The principal objectives were to identify conserved amino acids that are potentially critical for IGRP function and conserved sequences within the *IGRP* gene promoter that may specify its restricted tissue distribution. In the process we have uncovered a major difference in gene structure and expression between rodent species.

EXPERIMENTAL PROCEDURES

Materials—[α-^{32}P]dATP (>3000 Ci mmol^{-1}) and [γ-^{32}P]ATP (>6000 Ci mmol^{-1}) were obtained from Amersham Pharmacia Biotech, and [^{3}H]acetic acid, sodium salt (>10 Ci mmol^{-1}), was obtained from ICN. All individually specified reagents were of analytical grade and purchased from Sigma.

General Cloning, DNA Isolation, and Sequencing Procedures—Plasmid DNA purification, subcloning, and restriction endonuclease analyses were performed by standard protocols (29). DNA fragments used for subcloning and labeling were isolated from agarose gels using either the Qiaex II gel extraction kit (Qiagen) or Quantum Prep spin columns (Bio-Rad). Zeta-probe membranes (Bio-Rad) were used for DNA hybridization analysis and both alkaline transfer and hybridization using the standard protocol were performed according to the manufacturer's instructions. DNA probes were labeled by random oligonucleotide priming with [α-^{32}P]dATP using the Stratagene Prime-It II random primer labeling kit according to the manufacturer's instructions (final specific radioactivity, 0.2–1.2 Ci.nmol^{-1}). DNA sequencing was performed using the U. S. Biochemical Corp. Sequenase kit or by automated sequencing using an ABI 377 DNA analyzer. All DNA sequences are numbered relative to the experimentally determined mouse *IGRP* gene transcription start site, designated +1 (28).

Isolation of Human IGRP Genomic Clones—An arrayed human PAC library[2] was screened using a ^{32}P-labeled 1028-bp cDNA probe isolated as a *Pst*I fragment from the pSV.SPORT 1B1 clone, which contained sequences from all 5 exons of mouse IGRP (21). Thirty PAC library filters (1,105,920 clones) were incubated overnight in 200 ml of 6× SSC ((1× SSC = 0.15 M NaCl and 0.015 M sodium citrate)), 0.5% SDS, 100 μg/ml salmon sperm DNA, and 100 ng of labeled probe (~500,000 dpm/ml), washed five times at moderate stringency, and exposed for ~2 h at −80 °C to Kodak X-Omat AR film. Specific hybridization signals were identified as positive replicates from the filter map, and the corresponding PAC clone was then used for large scale isolation of plasmid by cesium chloride gradient centrifugation (29). Two independent clones designated PAC 294 (~90 kbp) and PAC 299 (~110 kbp) were isolated. They were digested with a panel of restriction enzymes, and Southern blot analysis was performed with fragments corresponding to the 5′ or 3′ end of the mouse *IGRP* gene (28). These were, respectively, a *Sal*I-*Pvu*II fragment of a mouse IGRP cDNA clone (pSV.SPORT 1B1) containing exons 1, 2, and part of exon 3 (21) and an *Xba*I-*Xba*I fragment of the mouse *IGRP* gene, isolated from the pGEM-BAC 4.5 plasmid (28), containing exons 4 and 5 and the intervening intron sequence. The hybridization analysis indicated that both PACs contained the entire human *IGRP* gene, which was then isolated within two overlapping genomic sub-clones, a ~9-kbp *Xho*I-*Xho*I fragment containing the promoter and exons 1–4 and an ~6-kbp *Eco*RI-*Eco*RI fragment containing exons 3–5 and 3′-flanking sequence. The fragments were sub-cloned from PAC294 into pGEM7 (Promega, Madison, WI) and sequenced on both strands over their entire length. The identification of the exon/intron boundaries (Table I) and the sizes of the four introns in the *IGRP* gene (Fig. 1) were initially determined by comparison with the mouse *IGRP* gene sequence (28) and subsequently from the human cDNA sequence (see below).

Chromosome Mapping—Contiguous sequence data spanning ~10 kbp of the cloned human *IGRP* gene was analyzed by the CENSOR program (30) to identify and edit out repeated genomic sequences and used to search the unannotated high throughput-sequencing human genome data base. Two BACs were identified; one (AC069137) containing the full-length gene on a 13.3-kbp contig within a 195-kbp insert the other (AC90045) containing a series of noncontiguous sequences within a 190-kbp insert. The BAC sequences were again edited by CENSOR to remove the repeated sequence and compared with the data base to identify genes flanking human IGRP.

As an independent approach, two PCR primer sets within exon 2 (forward 5′-ATGTGTGAGAGACCAAGACCTAAG-3′ and reverse 5′-TGAAGTTTTAGCATCCTCACTC-3′) and exon 5 (forward 5′-AGAACCTCTGTGTCTAATGC-3′ and reverse 5′-GGTCTGTGCCTACTCTGTGG-3′) were used to analyze the Stanford radiation hybrid panel G3, which has been mapped with 1185 markers. Reactions (10 μl) were run for 35 cycles in PCR buffer (PerkinElmer Life Sciences) containing 3 mM $MgCl_2$ and 30 ng of template (95 °C × 15 s, 55 °C × 15 s, and 72 °C × 30 s) using *Taq* gold polymerase (Applied Biosystems) and a GeneAmp PCR system 9700 machine (Applied Biosystems).

Isolation of Rat IGRP Genomic Clones—A fragment of the rat *IGRP* gene was generated using rat genomic DNA (CLONTECH) as the template in a PCR reaction with the following primers: forward (5′-CGGAATTCCTCCACAGATGGTCAGCATCACATG-3′) and reverse (5′-CGGAATTCGGGGTCTCCAACATTGGACATAAAATTTAG-3′); *Eco*RI cloning sites are underlined. The primers were designed based on conserved sequences in the human and mouse *IGRP* genes present in the promoter and exon 1, respectively. The PCR fragment generated was cloned into the *Eco*RI site of pGEM7 (Promega), and the 237-bp IGRP insert was subsequently used as a labeled probe to screen a rat BAC library (Genome Systems, Inc. Gene Screening Custom Service). A single rat BAC clone, designated 67/L18, hybridized to the probe. The large scale isolation of 67/L18 BAC plasmid DNA was performed by standard cesium chloride centrifugation (29).

BAC 67/L18 contained the entire rat *IGRP* gene. Restriction enzyme analysis and Southern blotting were performed as with the human *IGRP* gene using labeled fragments representing either the 5′ or 3′ end of the rat *IGRP* gene. The fragment representing the 5′ end of the rat gene was the same as that used in the initial library screening (see above, this section). The fragment representing the 3′ end of the rat *IGRP* gene was generated using the 67/L18 BAC plasmid as the template in a PCR reaction with the following primers: forward (5′-GGAATTCTCACGAGTCCAGCAAAAGGCGTG-3′) and reverse (5′-CCCAA-

[2] Children's Hospital Oakland Research Institute BACPAC Resources Home Page.

GCTTGAGGCCTTTGAACACACTCCAGG-3′); the *Eco*RI- and *Hin*dIII-cloning sites are underlined. These primers represent IGRP exon 5 sequences that are conserved in the human and mouse *IGRP* genes. The 221-bp rat IGRP fragment PCR fragment generated was cloned into and subsequently released from the *Eco*RI-*Hin*dIII-digested pGEM7 vector.

Genomic DNA fragments that hybridized to these labeled probes were then subcloned into the pGEM7 or pSP72 plasmid vectors (Promega) for sequence analysis. The entire rat *IGRP* gene was isolated within two overlapping genomic sub-clones; a ~6-kbp *Kpn*I-*Kpn*I fragment contained the promoter and exons 1–3, whereas a ~6-kbp *Pst*I-*Pst*I fragment contained exons 3 and 5. The identification of the exon/intron boundaries (Table I) was determined by direct DNA sequencing of both DNA stands and comparison with the mouse *IGRP* gene sequence (28). The sizes of the three introns in the rat *IGRP* gene (Fig. 1) were calculated by direct sequencing (introns A and C) or estimated by PCR (intron B). In the latter case, the size of the intron was estimated using two separate primer pairs; the difference in the size of the PCR products was as expected. PCR reactions (100 μl) contained 100 pmol of each primer, 1× PCR buffer (PerkinElmer Life Sciences), 0.2 mM each dNTP, 1.5 mM $MgCl_2$, 200 ng of pGEM7 plasmid DNA containing the ~6-kbp *Kpn*I-*Kpn*I rat IGRP fragment as the template, and 5 units of AmpliTaq DNA polymerase (PerkinElmer Life Sciences). Reactions were run for 94 °C for 5 min and then for 30 cycles of 30 s at 94 °C, 30 s at 47 °C, and 2 min at 72 °C using an MJ Research MiniCycler. Products were analyzed by agarose gel electrophoresis.

Cell Culture—The pancreatic islet-derived cell lines, βTC3, Min6, RIN, INS1, HIT, and αTC1 as well as COS 7 cells were passaged as subconfluent cultures (8.5-cm dishes) in Dulbecco's modified Eagle's medium supplemented with 100 units/ml penicillin and 100 μg/ml streptomycin. INS-1 cultures contained in addition 50 μM mercaptoethanol. All cell cultures were also supplemented with 10% (v/v) fetal bovine serum except HIT cell cultures, which were supplemented with 2.5% (v/v) fetal bovine serum and 15% (v/v) horse serum.

Northern Blotting and RT-PCR Analyses—Wistar-Furth rats and Balb-c mice were obtained from Charles River and Jackson Laboratories, respectively. Animals were fed *ad libitum* and sacrificed by CO_2 asphyxiation. Islets were isolated by a modification of the collagenase digestion procedure of Lacy and Kostianovsky (31) and Guest *et al.* (32). Viable human islets derived from cadaveric donors were obtained through the Juvenile Diabetes Foundation International human islet consortium, shipped at room temperature in Connaught Medical Research Laboratory medium supplemented with 10% fetal bovine serum and 5.6 mM glucose (Life Technologies), and maintained in culture for 24–72 h in the same medium under a 5% CO_2 in air atmosphere before harvesting RNA.

Total RNA was prepared from tissues and cell lines using Trizol reagent (Life Technologies) and quantified in a fluorimetric assay by DNase-resistant SYBR green binding (RiboGreen kit; Molecular Probes, Eugene, OR). Northern blotting was performed after electrophoresis of samples (5 μg of total RNA) on denaturing formaldehyde gels and with commercially available blots prepared with 2-μg samples of poly(A)$^+$ mRNA from various human tissue sources (multiple tissue northern and multiple tissue northern 1; CLONTECH, Palo Alto, CA). Blots were hybridized for 16 h at 42 °C in 50% (v/v) formamide, 5× saline/sodium phosphate/EDTA, 5× Denhardt's reagent, and 50 μg/ml salmon testis DNA with a ^{32}P-radiolabeled randomly primed probe corresponding to the ORFs of human IGRP or rat G6Pase catalytic subunit (9). Blots were subsequently washed in 2× SSC, 0.05% SDS at room temperature for 30 min, then in 0.2× SSC, 0.1% SDS at 42 °C and visualized by PhosphorImager (Molecular Dynamics, Palo Alto, CA).

Reverse transcription of human, mouse, and rat tissue total RNA (300 ng) was performed with oligo(dT_{18}) or random nonamers as primers using Maloney murine leukemia virus reverse transcriptase at 42 °C with reagents provided in the Stratagene high fidelity RT kit. Comparative analysis of the expression of IGRP in various tissues of the mouse, human, and rat was performed using a conserved primer pair within exons 1 (forward 5′-CCAAGATGAT(A/C)TGGGTAGC-3′) and exon 5 (reverse 5′-TGTCAATGTGGATCCAGTC-3′). The forward primer had a single degenerate nucleotide (internal parentheses (A/C)) to accommodate a single base difference between the mouse and human/rat sequences. PCR reactions were performed with a *Pfu*/*Taq* polymerase mixture (Roche Molecular Biochemicals Expand High Fidelity PCR system) for 5 min at 94 °C, then 40 cycles of 1 min at 94 °C, 1 min at 53 °C, and 2 min at 72 °C followed by a final 20 min extension at 72 °C. Products were analyzed on 1.5% agarose gels, and the major bands were excised and cloned into the pTOPO PCR blunt II vector using a zero blunt TOPO PCR cloning kit (Invitrogen, Carlsbad, CA) and then sequenced.

Human islet cDNA library screening using the mouse IGRP ORF probe proved unsuccessful. However, once the genomic sequence and exon/intron boundaries of human IGRP were established (Table I), each exon was amplified using a series of PCR forward and reverse primers, each incorporating 10 bp of the 5′ sequence of the flanking exon and the first 20 bp of the exon to be amplified. The products of these reactions were gel-purified, mixed, and subjected to PCR using a forward primer incorporating the start codon and Kozak sequence (33) (5′-TCAAGATGGATTTCCACAGGA-3′) and a compatible reverse primer located just beyond the stop codon (5′-CAGAGCACTAACTCTAGGCACC-3′). The sequence of the PCR product generated was confirmed and gave the expected *in vitro* translation product. Human cadaveric islet RNA became available at a later date and was used to amplify the coding region of human IGRP cDNA with 40 cycles of RT-PCR with the above forward primer and a reverse primer located further downstream in exon 5 (5′-GTGAAGTCGGATTAGAAGCC-3′). The PCR products were inserted into the pTOPO blunt vector and then subcloned into pCDNA3.1 for expression studies using *Eco*RI and *Xho*I sites common to both vectors.

Generation of Antisera and Immunoperoxidase Staining—A *Pst*I fragment containing the majority of the mouse IGRP ORF was inserted in-frame in the pUEX vector (34), generating a fusion protein with β-galactosidase, which was purified by isolation of inclusion bodies and preparative SDS-gel electrophoresis. Antibodies were raised in New Zealand white rabbits by immunization in complete Freund's adjuvant followed by boosting in incomplete Freund's adjuvant at six weekly intervals. Balbc mouse pancreas was perfusion-fixed with 4%(w/v) paraformaldehyde and subjected to standard paraffin embedding and sectioning before immunoperoxidase staining using the primary antiserum diluted 1:100 in PBS (2 h at room temperature) and a secondary donkey anti-rabbit antibody conjugated to horseradish peroxidase (Jackson Laboratories).

Fusion Gene Plasmid Construction and Analysis—The construction of a mouse IGRP-chloramphenicol acetyltransferase (CAT) fusion gene containing promoter sequence from −306 to +3 in the pCAT(An) expression vector (35) has been previously described (28). A rat IGRP-CAT fusion gene was constructed in the pCAT(An) expression vector as follows. The rat *IGRP* gene promoter was isolated as a *Hin*dIII-*Pst*I fragment and subcloned into *Hin*dIII-*Pst*I-digested pSP72 (Promega). The promoter fragment was then isolated from the pSP72 plasmid as a *Hin*dIII-*Bam*HI fragment and ligated into *Hin*dIII-*Bgl*II-digested pCAT(An). The resulting plasmid contains rat IGRP promoter sequence from −900 to +3, relative to the position of the mouse transcription start site. The *Pst*I site used in this cloning is conserved in the mouse *IGRP* gene, and the same strategy was used in the construction of the previously described full-length mouse IGRP-CAT fusion gene (28). Therefore, the same 3′ polylinker sequence between position +3 and the CAT reporter gene is present in the mouse and rat fusion gene constructs. A truncated rat IGRP-CAT fusion gene was then generated by restriction enzyme digestion of the −900 IGRP-CAT construct with *Hin*dIII and *Nhe*I followed by Klenow treatment of the noncompatible ends and blunt-end ligation. The resulting plasmid has a calculated 5′ end point of −321. The TGTA sequence in the rat IGRP promoter was mutated to a TATA box by site-directed mutagenesis within the context of the −321 to +3 promoter fragment using PCR and the following oligonucleotide as the 3′ primer: 5′-AACTGCAGGGCTCAGAGTTCGGTTGTCTTTATAGGGTCCCCTTGTGATG-3′. A *Pst*I site used for cloning purposes and the mutated base are underlined. The 5′ PCR primer (5′-CGGGATCC*AAGCTCTAGC*CAAGC-3′), with the *Bam*HI cloning site underlined, was designed to conserve the junction between the IGRP promoter and pCAT(An) vector to be the same as that in the wild-type rat − 321 IGRP-CAT fusion gene construct; the *Hin*dIII-*Nhe*I junction is shown in italics. The PCR fragment was digested with *Bam*HI and *Pst*I and subcloned into *Bam*HI-*Pst*I-digested pSP72 for sequencing. The promoter fragment was then re-isolated from the pSP72 plasmid as a *Bam*HI-*Pst*I fragment and ligated into *Bam*HI-*Pst*I-digested −321 rat IGRP-CAT. This *Bam*HI site is located immediately 5′ of the *Hin*dIII cloning site in the pCAT(An) vector (35).

A human IGRP-CAT fusion gene was constructed in the pCAT(An) expression vector such that the 5′ and 3′ end points were equivalent to those in the mouse −306 IGRP-CAT and rat −321 IGRP-CAT constructs (Fig. 7). This was achieved using PCR in conjunction with the following 5′ (5′-CCCAAGCTTCACCAAACATAGAAATTGC-3′) and 3′ (5′-AACTGCAGTGCTCTGATTCCCACCG-3′) primers. *Hin*dIII and *Pst*I sites used for cloning purposes are underlined. A single base pair change (italics) at position −1 in the human promoter was introduced

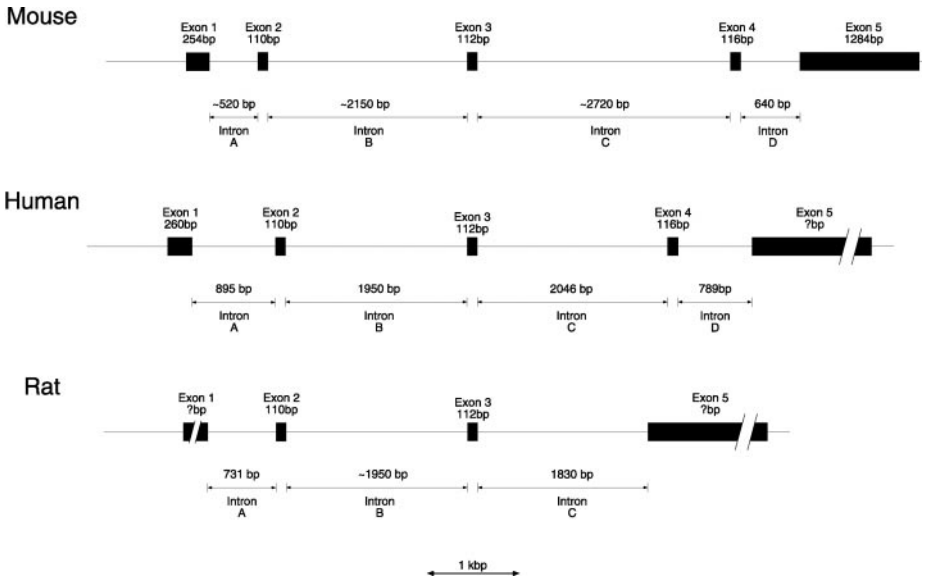

FIG. 1. **Structure of the mouse, rat, and human *IGRP* genes.** The *IGRP* gene exon and intron sizes were determined by a combination of direct DNA sequencing and PCR as described under "Experimental Procedures."

into the 3′ primer to restore the *Pst*I site such that the subsequent sub-cloning of the PCR fragment generated a fusion gene construct with the same 3′ polylinker sequence between position +3 and the CAT reporter gene as found in the mouse and rat fusion gene constructs. Thus, the PCR fragment was digested with *Hin*dIII and *Pst*I and subcloned into *Hin*dIII-*Pst*I-digested pSP72. The promoter fragment was then isolated from the pSP72 plasmid as a *Hin*dIII-*Bam*HI fragment and ligated into *Hin*dIII-*Bgl*II-digested pCAT(An). The resulting plasmid contains human IGRP promoter sequence from −324 to +3, relative to the position of the mouse transcription start site. Promoter fragments generated by PCR were completely sequenced to ensure the absence of polymerase errors, whereas promoter fragments generated by restriction enzyme digestion were only sequenced to confirm the 5′ end points. All plasmid constructs were purified by centrifugation through cesium chloride gradients (29). For fusion gene analyses, HIT cells were grown and co-transfected as previously described using a calcium phosphate precipitate containing 15 μg of a CAT construct and 2.5 μg of a Rous sarcoma virus-β galactosidase fusion gene construct (36). After transfection, β-galactosidase and CAT activity were assayed as described (36). To correct for variations in transfection efficiency, the results are expressed as a ratio of CAT:β-galactosidase activity. In addition, three independent preparations of each plasmid construct were analyzed in quadruplicate in separate experiments.

IGRP Protein Expression by in Vitro Translation and Cellular Transfection—In vitro transcription/translation assays were performed using rabbit reticulocyte lysate with a TNT T7 Quick translation kit (Promega) as previously described (37) using T7 polymerase transcripts from sequences cloned into the mammalian expression vector pCDNA3.1 (Invitrogen). A number of mouse IGRP constructs were analyzed in addition to the above-mentioned human IGRP cDNA clones to evaluate the effects of 5′- and 3′-untranslated region sequences on expression levels and the activity of two alternative start codons in the sequence. These included: (i) the full-length mouse IGRP cDNA (nt 1–1901) (21); (ii) a *Pst*I fragment (nt 110–1137) incorporating the second and third AUG codons; (iii) a cloned PCR product (nt 220–1038) generated using the primers 5′-TTGGAACCAAGATGATCTGG-3′ (forward) and 5′-CAGAGCACTAACTCTAGGCACC-3′) (reverse), which deleted the putative start codon and second AUG codon but retained the third potential AUG (nt 231) embedded in a Kozak sequence; (iv) a cloned PCR product (nt 59–1038) generated using the primers 5′-CAAGATGGATTTCCTTCATAGGAGT-3′ (forward) and 5′-CAGAGCACTAACTCTAGGCACC-3′) (reverse), which contains the entire ORF but with minimal flanking sequence.

The human IGRP protein was expressed by transient transfection of COS 7 cells using the pCDNA3.1 vector full-length construct (see above). A rat G6Pase catalytic subunit cDNA cloned into the same vector was used as a positive control (21). Transfections were performed as previously described using a calcium phosphate precipitate containing 15 μg of a pCDNA 3.1 construct and 5 μg of pRSV β-galactosidase followed by culture for a further 48–72 h in Dulbecco's modified Eagle's medium with serum (36). Cells were harvested using a non-enzymic procedure (Life Technologies, Inc. cell dissociation buffer), rinsed twice in 0.3 M sucrose, 10 mmol l^{-1} MES-K^+, 2 mmol l^{-1} EGTA, 1 mmol l^{-1} $MgSO_4$ (pH 6.5), and then sonicated for 20 s in 1 ml of the same media. The sonicate was centrifuged at 800 × *g* for 6 min to remove unbroken cells and debris, and a particulate fraction was prepared by further centrifugation of the supernatant at 214,000 × *g* for 30 min in a Beckman TLN-55 rotor. The pellet was resuspended in 300–500 μl of homogenization media (~0.3–1 mg/ml protein) and assayed for G6P hydrolytic activity (21). The supernatant was assayed for β-galactosidase activity using the spectrophotometric assay previously described (21).

RESULTS

Isolation of the Human and Rat IGRP Genes—The human *IGRP* gene was isolated from a human PAC library probed with a mouse IGRP cDNA fragment (Fig. 1). Two independent clones with similar restriction digestion patterns were obtained from a screen of ~10^{10} bp of genomic sequence. One of these clones, designated PAC 294, was selected for further analysis and was subsequently found to contain the entire human IGRP transcription unit (Fig. 1). A fragment of the rat *IGRP* gene was generated by PCR using primers representing a region of the promoter and exon 1 conserved in the mouse and human *IGRP* genes. This was used to probe a rat BAC library to obtain a single positive clone that was found to contain the entire rat *IGRP* gene (Fig. 1).

Exon/Intron Structure of the Human and Rat IGRP Genes—The exon/intron structure of the human *IGRP* gene (Fig. 1; Table I) was initially determined by comparing the sequence of the human *IGRP* gene with that of the mouse IGRP cDNA (21) and gene (28) and confirmed by subsequent sequence analysis of human IGRP cDNAs generated by RT-PCR from human islet RNA. The human and mouse *IGRP* genes are both composed of 5 exons, and the sizes of exons 2, 3, and 4 are identical (Fig. 1). The exon/intron splice junctions are also well conserved in comparison with the mouse gene and match the splice consensus sequence (38), with the exception of the boundary between the 3′ end of intron C and the 5′ end of exon 4 (Table I). Both the human and mouse *IGRP* genes exhibit a change in the nucleotide at the 5′ end of exon 4 from the consensus G to an A (Table I). This change may explain the frequent removal of exon 4 by differential splicing of the mouse (21) and human IGRP mRNA (see below). The TATA motif identified in the mouse *IGRP* gene promoter (28) is also conserved in the human IGRP promoter (see below). Since this motif is critical for determining the location of the transcription start site (39), we would predict that the *IGRP* gene transcription start site is identical in the mouse and human genes. If correct, exon 1 in the human gene will be 6 bp larger than in the mouse gene due to an insertion in the 5′-untranslated leader sequence (Fig. 1).

TABLE I

Comparison of the exon/intron boundaries of the mouse, rat, and human IGRP genes

The mouse IGRP exon/intron boundaries are from Ebert *et al.* (28); the human and rat IGRP gene exon/intron boundaries were determined as described under "Experimental Procedures." Regions of sequence divergence from the mouse IGRP gene are underlined (uppercase exon sequence) or are in bold type (lowercase intron sequence). The 5′ and 3′ consensus splice sequences are from Jackson (38).

Intron	Gene	5′ Intron junction	3′ Intron junction
A	Mouse IGRP	TTTAAATG/gtaagact	actcacag/GATATTGT
	Human IGRP	TTTAAATG/gtaaga**t**t	a**aat**acag/GATATTAT
	Rat IGRP	TTTAAATG/gtaagact	**gt**tcacag/GATATTGT
B	Mouse IGRP	AGGCCCAG/gtaagcag	cattgcag/GAAGTCCA
	Human IGRP	AGGTCCAG/gtaagc**ta**	**tg**ttgcag/GAAGTCCA
	Rat IGRP	AGGTCCAG/gtaagca**a**	ca**at**acag/GAAGTCCA
C	Mouse IGRP	CTGCACAG/gtcagctt	catcacag/ACTGACCT
	Human IGRP	CTGCACAG/gtcagctt	catc**gt**ag/ACTGACCT
	Rat IGRP	CTGCACAG/gtcagctt	Exon 4 absent
D	Mouse IGRP	GATTGGTG/gtaaatat	atccccag/GGATGCTA
	Human IGRP	AATTGGTG/gtaaatat	atcc**t**cag/GCATGCTG
	Rat IGRP	Exon 4 absent	atccc**t**ag/GGACGCTA
	Consensus sequence:	(A or T)G/gtaa	cag/G 3′

Thus, the length of IGRP-coding sequence in exon 1 (218 bp) is identical between human and mouse genes.

The exact size of human IGRP exon 5 is unknown since a human poly(A)$^+$ cDNA was not isolated. However, the length of the IGRP-coding sequence in exon 5 (512 bp) is identical in the human and mouse genes, and the human genomic sequence could be aligned with the mouse cDNA through to a conserved element preceding a consensus poly(A) addition site in mouse IGRP. The human IGRP genomic sequence up to this point contained an additional 500 bp appearing as separate 400- and 100-bp inserts. On this basis, the expected human IGRP mRNA would be larger than mouse IGRP mRNA, which is consistent with what is seen on Northern blots (see below). The four introns in the human *IGRP* gene, which were determined by direct sequencing, were similar in size to those of the mouse gene (Fig. 1).

The rat *IGRP* gene, by contrast, showed major differences from the mouse and human genes. The exon/intron structure (Fig. 1; Table I), determined by comparison of the rat *IGRP* gene and mouse IGRP cDNA (21) and gene sequences (28), showed that although exons 2 and 3 are identical in size, exon 4 is absent in the rat gene (Fig. 1). With the exception of exon 4, the exon/intron splice junctions are otherwise conserved (Table I), and the equivalent of exon 5 was identifiable by sequence homology. Direct sequencing showed that the intervening sequence between exons 3 and "5" in the rat *IGRP* gene was 1830 bp in length; the equivalent in the mouse and human was 3476 and 2951 bp, respectively. Alignment of the rat and human gene sequences indicated that ~500 bp were deleted both upstream and downstream of exon 4 (116 bp in both human and mouse). Introns A and B were of a similar size to the corresponding human and mouse introns (Fig. 1). The absence of exon 4 is consistent with other observations (see below) that indicate that the rat gene is a non-expressed pseudogene. The sizes for exons 1 and "5" in the rat gene thus cannot be assigned (Fig. 1).

Chromosomal Mapping of the Human IGRP Gene—Analysis of the human genome high throughput sequence data base identified two BACs of approximately 180 kbp, one of which contained the human *IGRP* gene as a contiguous sequence. The two BACS contained the human Unigene expressed sequence tag cluster markers H210260 and H101282, which placed the gene on the distal end of chromosome 2 in the interval D2S156 (microsatellite AFM211yd6) to D2S376 (microsatellite AFM319 × *g*1) (NCBI GeneMap 99 170.5–180.6 centimorgan) and close to D2S399 at 174.8 centimorgan (microsatellite AFMa131wb9). Radiation hybrid analysis placed the gene adjacent to the STS marker SHGC 13934 on chromosome 2 (LOD score 11.84 and 14.90 with exon 2 and exon 5 probes, respectively), which again lies in the D2S156 to D2S376 interval. The SHGC 13934 STS marker amplifies a 137-bp fragment of the myosin heavy chain 1B gene, the sequence of which was found within the same BAC as the human *IGRP* gene. Our previous mapping studies with mouse *IGRP* gene (*G6pc-rs)* using two interspecific back-cross DNA mapping panels located it on the proximal portion of mouse chromosome 2 near the marker *D2Mit11*, positioned at 39 centimorgan (28). The orthologous gene in humans would be on chromosome 2q, consistent with our observations.

Sequence Analysis and Translation of Human IGRP mRNA—Attempts to clone a cDNA for human IGRP from a series of human pancreatic islet cDNA libraries using a 1000-bp ORF probe from the mouse IGRP cDNA were unsuccessful. cDNAs encoding the ORF, however, could be generated either by PCR-based ligation of the individual exons or by RT-PCR from cadaveric human pancreatic islet total RNA. The synthetic and RT-PCR constructs were identical in sequence and produced the same sized products upon *in vitro* translation.

In vitro translation of a cRNA that incorporated the deduced ORF (nt 43–1020 relative to the mouse cDNA sequence) (21) generated a protein doublet of 37 kDa, a size consistent with the predicted molecular mass (40,583 Da) (Fig. 2). *In vitro* translation of the ORF of rat G6Pase catalytic subunit from the same vector gave a somewhat smaller product (33 kDa *versus* 40,559 Da, predicted) and raised the question as to whether the assigned human IGRP start codon was correct. Exon 1 of human IGRP has an in-frame stop codon at position 21, indicating that translation initiation upstream was not possible. Downstream, however, there are two start codons that are conserved in human IGRP and mouse IGRP, although not in the human or mouse G6Pase catalytic subunits. The second of these has a strong predicted Kozak sequence (33) that produced a protein of 28 kDa (predicted size 33.8 kDa) from a 5′-truncated construct (Fig. 2). It was concluded that the AUG codon at nt 43 is the preferred start site. The size discrepancy between the predicted and observed molecular mass of human IGRP probably relates to the hydrophobic nature of the protein. The slower electrophoretic mobility of IGRP *versus* the G6Pase catalytic subunit is conceivably related to its more acidic nature (predicted pI 8.72 *versus* 9.22). The molecular sizes of *in vitro* translated mouse IGRP and human IGRP were indistinguishable and consistent with the native predicted molecular mass (40,685 Da) and pI (8.62) of mouse IGRP and its size, deter-

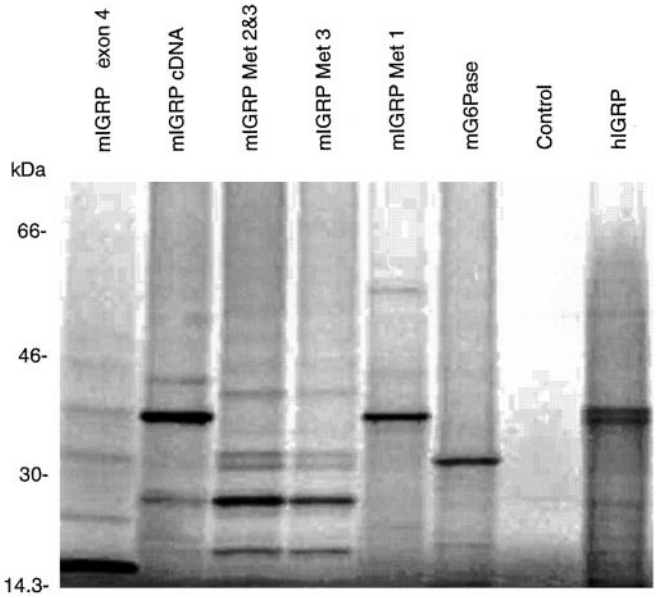

FIG. 2. ***In vitro* translation of human (*h*) and mouse (*m*) IGRP.** T7 polymerase-derived transcripts from cloned pCDNA3.1 constructs were translated *in vitro* in the presence of [^{35}S]methionine using rabbit reticulocyte lysate and analyzed by SDS-PAGE. Mouse IGRP constructs included the original cDNA (*mIGRP cDNA*; nt 1–1901), a *Pst*I fragment containing the second and third alternative start codons (*mIGRP Met 2&3*; nt 110–1137), a PCR construct containing the third start codon only (*mIGRP Met 3*; nt 220–1038), a PCR construct containing the ORF but with minimal flanking sequence (*mIGRP Met1*; nt 59–1038), and the Δexon 4 form of IGRP (*mIGRP exon 4*).

mined by Western blot analysis of islets (38 kDa) (data not shown).

The deduced human IGRP ORF encoded a 355-amino acid protein of a generally hydrophobic character (Fig. 3). Three consensus sites for NH_2-linked glycosylation were present (amino acids 50, 92, and 287), and the protein had a COOH-terminal consensus sequence (KK*XX*) characteristic of an ER-resident transmembrane protein. The hydrophobic amino acids were arranged in nine major stretches, eight of which were predicted to be able to span a phospholipid bilayer as an α-helix (TMAP (40)). Each stretch, however, contained a charged amino acid(s) (aa): the sequence aa 25–47, Asp^{34} and Arg^{36}; aa 57–77, Asp^{65} and Lys^{72}; aa 116–138, His^{133}; aa 148–173, Arg^{168}; aa 179–193, Glu^{191}, aa 210–230, Arg^{227}; aa 255–280, Arg^{261} and Glu^{276}; aa 288–307, Arg^{293}; and residues 319–343, Lys^{327}. The NH_2 terminus of the protein did not bear a consensus signal sequence for import into the ER, but it is conceivable that the putative transmembrane segments could function in this regard.

The human IGRP ORF sequence could be aligned with 2 gaps with residues 7–359 of the human G6Pase catalytic subunit (359 aa) (Fig. 3). The sequence was 50.4% identical (75.2% similarity) and homologous over the full length of the molecules including the putative transmembrane domains. The conservation of these hydrophobic segments and the charged residues within them suggests that they may have a function other than simple membrane spanning. Of the three potential sites of *N*-glycosylation, only the site located in the putative second lumenal domain (amino acid 92) was conserved. Both the human IGRP and G6Pase catalytic subunit molecules contain the COOH-terminal ER membrane protein retention motif. The human IGRP ORF was identical in length to the previously cloned mouse IGRP (21) and was closely homologous (84.8% identity; 92.9% similarity). The homology is similar to that between the human and mouse G6Pase catalytic subunits and argues for conservation of function of IGRP. The G6Pase catalytic subunit shares an extended sequence motif that is found in bacterial vanadate-sensitive haloperoxidases and mammalian phosphatidic acid phosphatases (41, 42) (Fig. 3). This motif, which incorporates the active site of these enzymes, is conserved in human IGRP (Lys^{72} . . . Arg-Pro^{80} . . . Pro-Ser-Gly-His^{115} . . . Ser-Arg^{168} . . . His^{174}) with the same alignment. Amino acids within the G6Pase catalytic subunit sequence whose mutation results in loss of function in glycogen storage disease type 1a (Fig. 3) were generally conserved with the exception of Gly^{33} (Ala), Ser^{120} (Ala), and Lys^{209} (Leu); the amino acid present in the G6Pase catalytic subunit is shown in parentheses. Human and mouse IGRP were identical in sequence at all these positions.

No cDNA for rat IGRP could be isolated by screening either rat insulinoma or rat islet cDNA libraries with a mouse IGRP probe nor generated by RT-PCR of rat islet total RNA using rat-specific primers deduced from the genomic sequence. This contrasted with the relative ease with which the mouse IGRP cDNA was obtained from these sources (21). The alignment of the coding regions of the rat gene and those of mouse and human IGRP furthermore suggested that even if such a cDNA existed, it would not encode a protein of similar size and sequence to mouse or human IGRP. Thus, although regions of the rat *IGRP* gene corresponding to exons 1, 2, 3, and 5 of mouse IGRP were highly homologous at the nucleotide level (89.9% identity), there were 3 catastrophic changes within the deduced reading frame. First, a deletion of 2 nucleotides within exon 1 (mouse nt 73 and 89) changes the rat coding sequence beyond amino acid 10 even though the reading frame remains open to the end of the exon. Second, exon 4 (116 bp) is absent, and splicing of exons 3 and 5, if it occurred, would alter the reading frame of exon 5. Third, an additional base is present in exon 5 (mouse nt 917), altering the reading frame and producing a premature stop codon. The change in sequence of exon 1 would be circumvented if the alternate start site at Met^{57} were used; however, this would only generate a 11.8-kDa protein because of the exon 4 deletion.

Tissue Distribution and Expression of Human IGRP mRNA—Northern blot analyses of human tissue poly(A)$^+$ mRNA with a human IGRP ORF probe showed the presence of a single ~3100-bp hybridizing species in pancreas (Fig. 4*A*). Testis produced a weaker signal (10% of the pancreas signal) from hybridizing species of ~2400 and 1000 bp (Fig. 4*B*), whereas 14 other major human tissues were negative (<3% of the pancreas signal). The same probe used under low stringency conditions (42 °C, 0.2× SSC) showed a strong signal from human pancreatic islet total RNA, an even stronger signal with the equivalent loading of mouse islet RNA, but no signal from rat islets or the rat insulinoma cell lines, RIN (Fig. 4*C*) and INS-1 (data not shown).

IGRP mRNA expression was further investigated by RT-PCR using highly conserved primer pairs within exons 1 and 5. As in the case of Northern blotting, strong signals were detected from mouse and human islet RNA preparations, but none were detected from rat islets (Fig. 4*D*). A series of RT-PCR products that were ~100, 200, and 350 bp shorter than the expected 595-bp target were also obtained using RNA from human islets and human β cell-derived cell lines. Cloning and sequencing of these products from human sample 144 and 1174 (Fig. 4*D*) identified them as alternatively spliced variants, the most prominent of which is an exon 4 deletion equivalent to that previously documented in mouse IGRP (21, 28). Other variants included deletions of exon 2, exons 3 plus 4, and exons 2, 3, and 4 together. Splicing occurred accurately at the donor/splice junctions shown in Table I. Only one of the alternatively spliced products (Δexons 3 and 4) maintained the reading frame of the full-length molecule and could potentially gener-

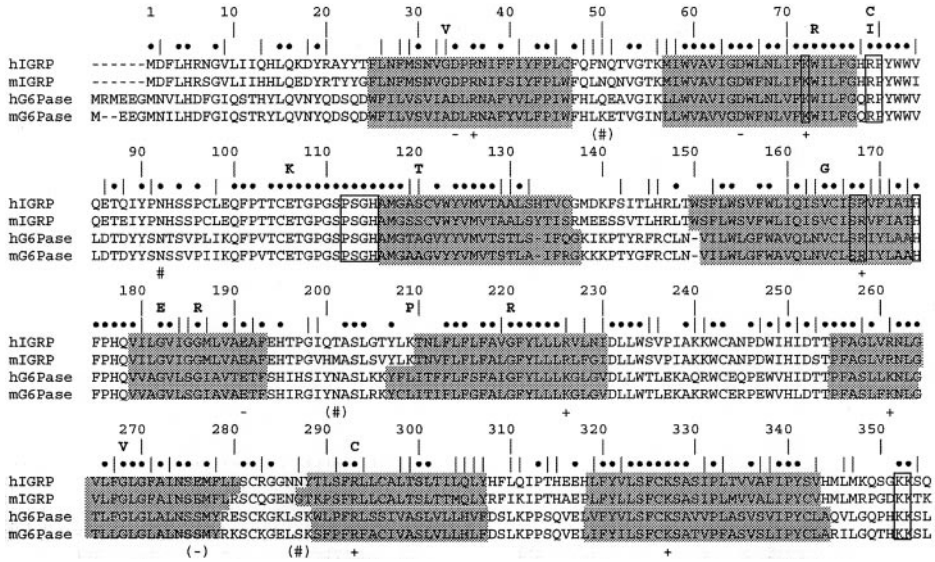

FIG. 3. **Alignment of the deduced peptide sequences of human and mouse IGRP and the corresponding G6Pase catalytic subunit sequences.** The predicted amino acid sequence of the human IGRP protein was aligned using CLUSTAL with mouse IGRP, human G6Pase catalytic subunit (accession number U01120), and mouse G6Pase catalytic subunit (accession number U00445). Putative transmembrane segments are *shaded*, and conserved charged residues within them are designated +/−. Consensus sites for NH_2-linked glycosylation are also shown (#) below the sequence block. Residues defined as being of key catalytic importance in haloperoxidases and related phosphatases are *boxed*, as is the COOH-terminal endoplasmic reticulum retention signal. Point mutations in the human liver enzyme that give rise to type 1a glycogen storage disease are indicated by the depiction of the mutant residue above the sequence block. Black dots (●) indicate amino acids that are identical between the human and mouse IGRP and the human and mouse G6Pase catalytic subunit.

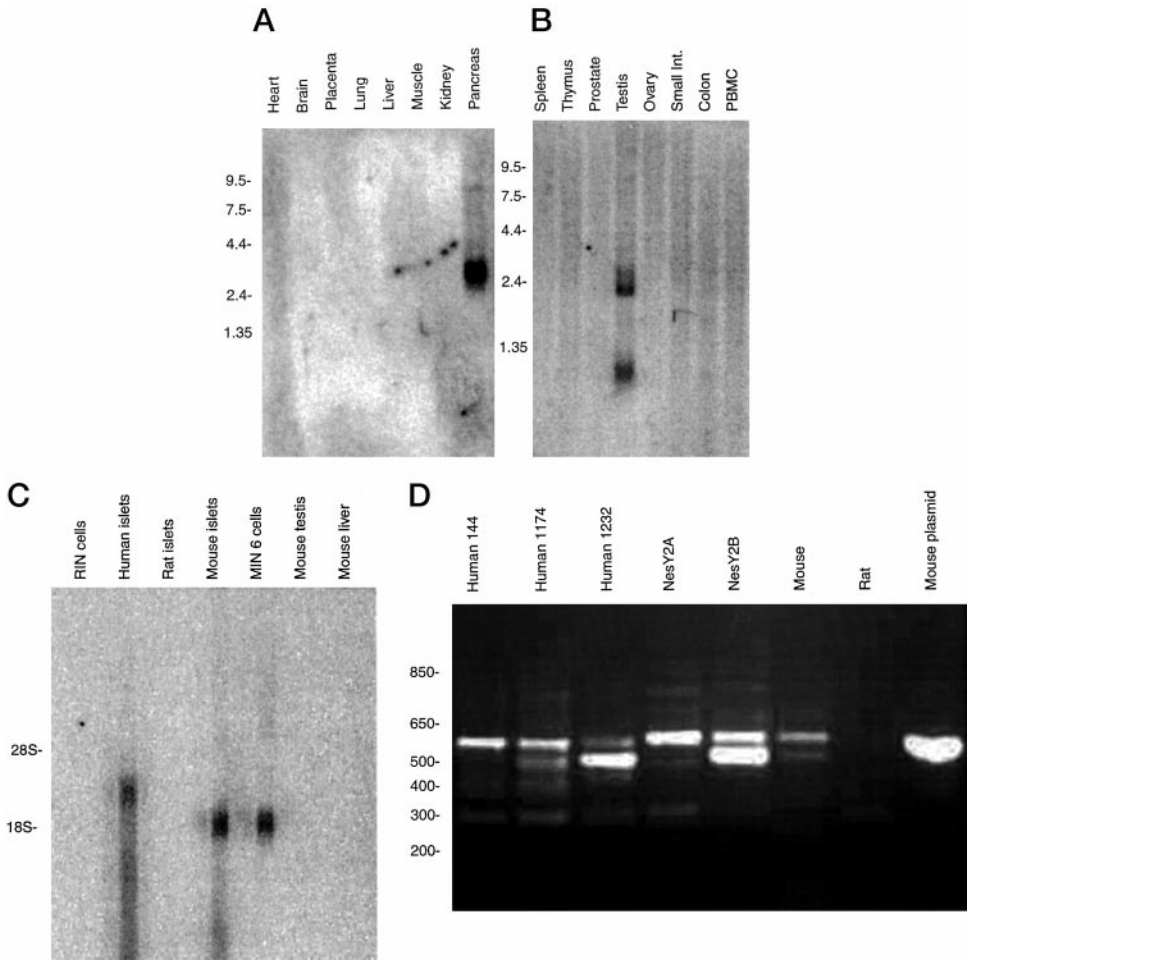

FIG. 4. **Analysis of mouse, rat, and human IGRP mRNA expression by Northern blotting and RT-PCR.** *Panels A* and *B*, Northern blotting analysis was performed on 2-μg samples of poly(A)$^+$ mRNA (CLONTECH MTN and MTN1 blots) hybridized with a human IGRP cDNA spanning the ORF. Blots were washed in 2× SSC at 42 °C and visualized by phosphorimaging (72 h exposure). The imaging sensitivity range is 3-fold higher for the blot shown in *panel B* relative to that in *panel A*. *PBMC*, peripheral blood mononuclear cell. *Panel C*, Northern blotting analysis was performed on 5-μg samples of total RNA from the indicated tissues hybridized with a human IGRP cDNA spanning the ORF. Blots were washed in 2× SSC at 42 °C and visualized by phosphorimaging (72-h exposure). *Panel D*, samples (300 ng) of total RNA were reverse-transcribed using random nonamers, then amplified by PCR for 30 cycles using conserved primers within IGRP exons 1 and 5. A full-length mouse IGRP clone (0.1 ng) was used as a reference template. Products included the expected 595-bp product and a series of shorter amplicons. Human islets were obtained from cadaveric donors (numbers 144, 1174, and 1232) and cultured for 24–72 h before analysis. Mouse and rat islets were freshly prepared. RNA isolated from two human β cell lines (NesY2 (human nesidioblastoma-derived insulin-secreting cell line) A and B) derived from a single nesidioblastoma patient was also analyzed.

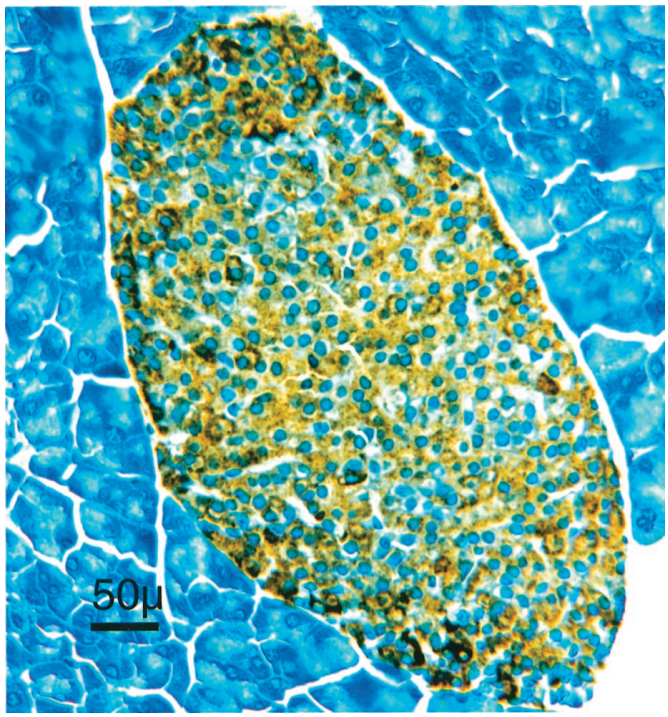

FIG. 5. **Immunoperoxidase staining of mouse pancreas with antibodies raised to recombinant IGRP.** Immunoperoxidase staining was performed as described under "Experimental Procedures." IGRP was localized by immunoperoxidase labeling (*brown*) and the sections were counterstained with Gill's hematoxylin.

ate a 32.2-kDa protein. The relative proportions of alternately spliced forms varied from individual to individual and between oligo(dT) *versus* randomly primed transcripts. This variation did not appear to correlate with the age, sex, and weight of the donor nor the warm ischemia time before islet isolation or duration of tissue culture before RNA extraction (Fig. 4*D*). Without reference to fresh material from healthy individuals, the significance of these alternatively spliced products is unclear.

Islet-specific Pancreatic Expression of Mouse IGRP—Although Northern blotting indicated a pancreas-specific pattern of IGRP mRNA expression (Fig. 4*A*), this analysis did not determine whether IGRP was expressed in pancreatic endocrine or exocrine cells. To address this issue antibodies were raised to recombinant mouse IGRP, and immunoperoxidase staining of mouse pancreas was performed. The result shows that the antigen was localized to islet cells with no reactivity evident in the acinar tissue or ductal elements (Fig. 5). Very few IGRP-negative cells were observed within the islet, suggesting that alpha and beta cells were certainly immunoreactive and that possibly all four endocrine cell types expressed the protein (Fig. 5).

Enzymic Activity of Human IGRP—Enzyme activity studies were performed by transiently transfecting COS 7 cells with various pCDNA 3.1 constructs; G6P hydrolytic activity was then assessed in a micosomal fraction prepared from lysed cells. A construct encoding the rat G6Pase catalytic subunit served as a positive control, and the efficiency of transfection was evaluated by co-transfection of a Rous sarcoma virus-β galactosidase fusion gene construct. Transfection with the G6Pase catalytic subunit construct resulted in an ~25-fold increase in G6P hydrolysis over basal activity (Table II). In contrast, transfection with a construct encoding human IGRP produced no detectable change in G6P hydrolytic activity, as previously observed for mouse IGRP (Table II and Ref. 21). Transfection of COS 7 cells with constructs encoding truncated

TABLE II

G6P hydrolytic activity in transiently transfected COS7 cells

COS7 cells were co-transfected as described under "Experimental Procedures," with pCDNA3.1 constructs encoding IGRP or rat G6Pase catalytic subunit (15 μg) together with a reference vector encoding β-galactosidase (5 μg). Enzyme activities were determined in cell homogenates after 48 h. Mouse IGRP constructs (see Fig. 2 legend for details) included the designated full-length construct (mouse IGRP), a *Pst*I fragment containing the second and third alternative start codons (Met 2+3), and a PCR construct containing the third start codon only (Met 3). Each tabulated result represents the mean ± S.E. of duplicate determinations from four separate experiments.

	β-Galactosidase	G6P hydrolysis
	nmol/min/mg	*nmol/min/mg*
Control	307 ± 103	6.1 ± 2.1
Human IGRP	239 ± 126	5.7 ± 2.7
Mouse IGRP	309 ± 111	6.3 ± 1.8
Met 2+3	400 ± 240	6.3 ± 3.1
Met 3	256 ± 148	6.6 ± 2.7
G6Pase	204 ± 91	152.6 ± 36.6

forms of mouse IGRP in which the putative start codon was deleted but which contained either the second or third putative start site also failed to increase basal G6P hydrolytic activity (Table II). The rates of hydrolysis of the generic phosphatase substrate, *p*-nitrophenol phosphate were not altered in human IGRP- or mouse IGRP-transfected COS 7 cells, although they were good substrates for the G6Pase catalytic subunit (data not shown and Ref. 21).

Transcriptional Activity of the Proximal Mouse, Rat, and Human IGRP Gene Promoters—We have previously shown that the proximal mouse IGRP promoter region, located between −306 and +3, is sufficient to confer maximal IGRP-CAT fusion gene expression in HIT cells (28). The level of basal mouse IGRP-CAT fusion gene expression in both HIT and βTC-3 cells decreases gradually upon deletion of the IGRP promoter sequence between −306 and −66, indicating that multiple *cis*-acting elements contribute to maximal fusion gene expression (36). An alignment of the equivalent human and rat IGRP promoter regions revealed multiple regions of conserved sequence (Fig. 6). We previously determined the location of several transcription factor binding sites in the mouse IGRP promoter using the ligation-mediated PCR *in situ* footprinting technique; these binding sites correlated with regions of the IGRP promoter, identified as being important for basal IGRP-CAT fusion gene expression (36). Fig. 6 shows that many of the residues in the mouse IGRP promoter that are contacted by transcription factors in βTC-3 cells *in situ* are also conserved in the human and rat promoters. In addition, a hepatocyte nuclear factor-3 binding site identified in the mouse IGRP promoter, which binds a hepatocyte nuclear factor-3 *in vitro* (36), is also conserved in the rat and human promoters (Fig. 6).

The observation that several putative *cis*-acting elements are conserved in the rat IGRP promoter (Fig. 6) was surprising given that the rat *IGRP* gene is not expressed (Figs. 4, *C* and *D*). In contrast, the TATA box motif identified in the mouse IGRP promoter is not conserved in the rat promoter (Fig. 6). Wobbe and Struhl (43) have shown that the sequence TGTA found in the rat promoter directs a greater than 20-fold lower level of *in vitro* transcription than the TATA motif. To determine whether this and other sequence variations affect the relative activity of the mouse, rat, and human IGRP promoters, we constructed fusion genes in which these promoters were ligated to the CAT reporter gene. Basal IGRP-CAT fusion gene expression was then assayed after transient transfection of the HIT cell line. Fig. 7*A* shows that the human IGRP promoter sequence located between −324 and +3 confers a slightly higher level of basal fusion gene expression than the equivalent

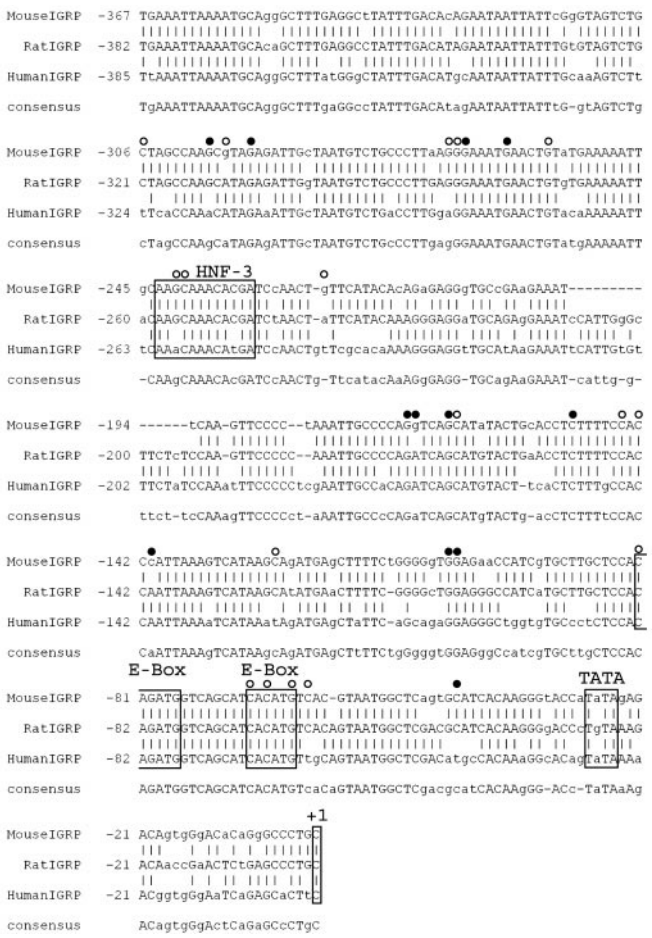

FIG. 6. **Alignment of the mouse, rat, and human *IGRP* gene promoter sequences.** The human, mouse, and rat IGRP promoter sequences were aligned using the IntelliGenetics, Inc. IFIND program and labeled relative to the experimentally determined transcription start site of mouse IGRP (28) designated as +1. Increases (●) or decreases (○) in dimethyl sulfate methylation of the mouse IGRP promoter comparing *in situ versus in vitro* methylated βTC-3 cell DNA were determined by ligation-mediated PCR (36). The TATA box, two E-box motifs, and a hepatocyte nuclear factor-3 binding site, which are conserved between the mouse and human promoters, are *boxed*.

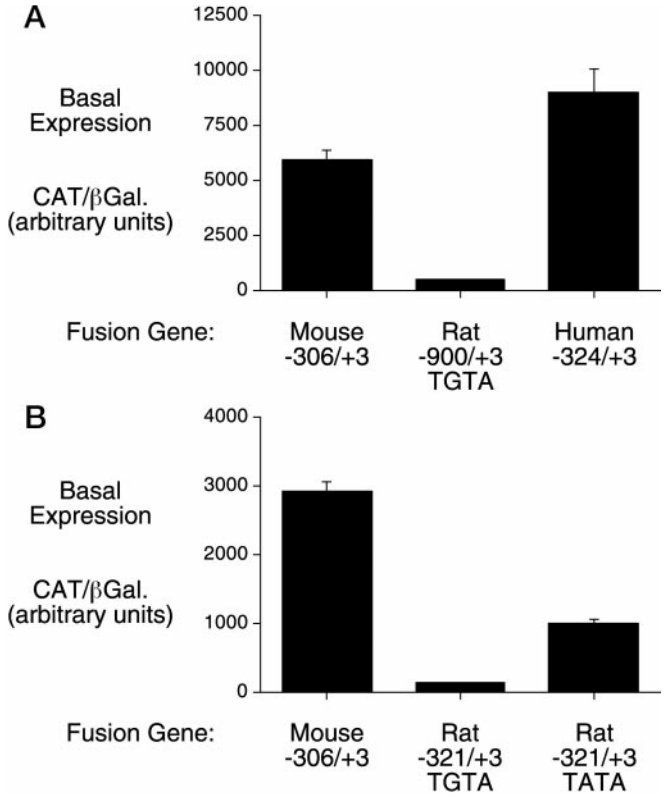

FIG. 7. **Basal activities of the mouse, rat, and human *IGRP* gene promoters.** HIT cells were transiently co-transfected, as described under "Experimental Procedures," with mouse, rat, or human IGRP promoter-CAT fusion gene constructs (15 μg) containing the promoter sequence shown together with a reference vector encoding β-galactosidase (*βgal*; 2.5 μg). The rat constructs had either the wild-type (TGTA) or back-mutated (TATA) box motif. After transfection, cells were cultured for 18–20 h in serum-free medium, and CAT and β-galactosidase activity was determined (36). The mean ratio of CAT:β-galactosidase activity ±S.E. is presented from three transfection experiments, each using independent preparations of each CAT plasmid.

mouse promoter sequence located between −306 and +3. By contrast, neither the equivalent rat IGRP promoter sequence located between −321 and +3 (Fig. 7*B*) nor a longer fragment of the promoter containing sequence located between −900 and +3 (Fig. 7*A*) confers appreciable fusion gene expression. However, mutating the rat TGTA motif back to the consensus TATA motif markedly enhances basal rat IGRP-CAT fusion gene expression (Fig. 7*B*). Nevertheless, the actual level of rat promoter activity remains ~3-fold lower than that of the mouse (Fig. 7*B*), suggesting that changes in elements other than the TATA box contribute to the low activity of the rat promoter.

A concern with all the data obtained on the rat *IGRP* gene and its transcription or translation was the fact that a single BAC clone served as the source of sequence data and fusion gene constructs. It was conceivable that a pseudogene was selected in the original screen and/or that sequence artifacts were introduced during the construction of the BAC library. These concerns were addressed by sequence analysis of critical regions of the rat *IGRP* gene amplified by PCR using primer sets based on conserved rat/mouse sequences with genomic DNA isolated from adult rat spleen and liver as the template. These regions included the promoter and the TGTA motif, the deletions and insertion affecting the reading frame in exons 1 and 5 and the entire intervening sequence between exons 3 and 5. In every case the BAC sequence was confirmed. The rat BAC sequencing and PCR-based gene sequencing were performed in Nashville, TN and Denver, CO, respectively using independent primer sets to avoid any risk of contamination.

DISCUSSION

Two remarkable features of the IGRP molecule are the focus of the current investigation. The first is its structural similarity to the G6Pase catalytic subunit, a molecule that plays a central role in glucose homeostasis and the pathophysiology of diabetes mellitus. The second is its restricted expression to pancreatic islets. A comparative analysis of the human, mouse, and rat gene structures was undertaken to define the conserved primary sequence of the protein, which could impact on the expression of IGRP catalytic activity and conserved promoter sequences that could be important in tissue-specific and physiological transcriptional regulation.

Multiple shared sequences were identified in the promoters of the mouse, rat, and human *IGRP* genes (Fig. 6) that, along with our previous *in situ* footprinting studies (36), will facilitate the identification of *cis*-acting elements, which are important for basal and islet-specific *IGRP* gene expression. Of specific interest for future mutagenesis are two putative E-box elements (Fig. 6). Such an element is one of the major sites that contributes to basal insulin gene expression mediated by a heterodimeric complex of the basic helix-loop-helix proteins BETA2/NeuroD and a ubiquitous factor, either E2A or HEB (44, 45). Surprisingly, both the putative IGRP E-box motifs are

conserved in the rat promoter despite the fact that the TATA box in the rat IGRP promoter is mutated (Fig. 6) and the gene is not expressed (Fig. 4*D*). An E-box motif in the human transcobalamin II promoter has been shown to mediate bidirectional transcription in the absence of a TATA box motif (46). However, in the case of the rat IGRP promoter, the two E-box motifs are not sufficient to confer high promoter activity in the absence of the TATA box (Fig. 7*B*). The proximal region of the mouse IGRP promoter that contains the two putative E-box motifs by itself confers very low basal fusion gene expression (36); thus, even if these elements are found to contribute to basal *IGRP* gene expression, their activity will depend on additional distal elements.

Several protein binding sites have been identified in the mouse IGRP promoter by *in situ* footprinting (36) (Fig. 6) for which no candidate *trans*-acting factor can be identified by computer analysis using the MatInspector software (47). These conceivably represent binding sites for novel islet-enriched *trans*-acting factors, a hypothesis that gains further credence from the observation that these regions are highly conserved in the rat and human IGRP promoters (Fig. 6). The search for such elements and their associated binding proteins is of particular interest given the emerging realization that such factors include potential diabetogenes (48, 49), many of which also play a critical role in pancreatic ontogeny (50–52).

Since the human *IGRP* gene, its mRNA, and the encoded protein showed great similarity to the corresponding mouse IGRP molecules this suggests conservation of biological function. The full-length human IGRP protein, like the mouse protein, possessed characteristic structural features of an ER-localized transmembrane protein with hydrolytic properties (41, 42). Yet as previously observed, no catalytic activity was obtained with either G6P or a generic phosphatase substrate (Table II). Nevertheless, the present documentation of the human IGRP protein sequence will assist in future structure/function studies using site-directed mutagenesis and chimeric IGRP/G6Pase catalytic subunit proteins. Immunohistochemical analyses with antibodies directed toward recombinant mouse IGRP confirm that the IGRP protein is expressed endogenously in mouse islets (Fig. 5). Moreover, Western blotting analyses with antibodies directed toward Myc-tagged mouse IGRP show that it can be expressed in COS 7 cells by transient transfection.[3] Thus our inability to detect enzymatic activity cannot be related to failure of protein expression.

Although the data obtained on human IGRP were consistent with the hypothesis that IGRP plays an important tissue-specific function in the adult pancreatic islet, the results with the rat *IGRP* gene paradoxically indicate that IGRP is a nonfunctional or perhaps a vestigial gene product. We found no evidence for the existence of a second, functional *IGRP* gene in the rat, and mRNA expression could not be detected in rat islets or rat islet-derived cell lines using either probes that hybridize with the highly homologous mouse sequence or by RT-PCR analyses with exact primers (Fig. 4). The rat *IGRP* gene nucleotide sequence was more closely related to the mouse than the human sequence, as expected from the taxonomic divergence of these species, and clearly it was the rat equivalent of the *IGRP* gene. A possible explanation of this paradox is that selection has occurred in the rat *IGRP* gene for a series of mutational events that have ensured its inactivation. The deletion of exon 4 in the rat *IGRP* gene removes sequences that would be essential for hydrolytic activity (53), although it could be argued that, like the common Δexon 4 splice variant of mouse (21) and human IGRP (Fig. 4D), such a molecule has a regulatory function. However, the reading frame mutations in the 5′ and 3′ ends of the ORF of rat IGRP and the change in the TATA box indicate that, at some point in evolution, expression of IGRP in the rat islet was turned off because the protein was unnecessary, redundant, or deleterious.

[3] B. Bergman and J. C. Hutton, unpublished observations.

Acknowledgments—We thank Roland Stein and Eva Henderson for providing the HIT cell line, Rebecca Taub for the rat G6Pase construct, Kevin Docherty for NesY2 (human nesidioblastoma-derived insulin-secreting cell line) RNA, and Donna Curtis for assistance with the data base searching. Human islets were obtained from Miami, FL, Rochester, MN and Boston, MA through the Juvenile Diabetes Foundation International islet consortium.

REFERENCES

1. Mithieux, G. (1997) *Eur. J. Endocrinol.* **136,** 137–145
2. Foster, J. D., Pederson, B. A., and Nordlie, R. C. (1997) *Proc. Soc. Exp. Biol. Med.* **215,** 314–332
3. Gerin, I., Veiga-da-Cunha, M., Achouri, Y., Collet, J. F., and Van Schaftingen, E. (1997) *FEBS Lett.* **419,** 235–238
4. Chou, J. Y., and Mansfield, B. C. (1999) *Trends Endocrinol. Metab.* **10,** 104–113
5. Cline, G. W., Rothman, D. L., Magnusson, I., Katz, L. D., and Shulman, G. I. (1994) *J. Clin. Invest.* **94,** 2369–2376
6. Consoli, A. (1992) *Diabetes Care* **15,** 430–441
7. Argaud, D., Zhang, Q., Pan, W., Maitra, S., Pilkis, S. J., and Lange, A. J. (1996) *Diabetes* **45,** 1563–1571
8. Liu, Z., Barrett, E. J., Dalkin, A. C., Zwart, A. D., and Chou, J. Y. (1994) *Biochem. Biophys. Res. Commun.* **205,** 680–686
9. Haber, B. A., Chin, S., Chuang, E., Buikhuisen, W., Naji, A., and Taub, R. (1995) *J. Clin. Invest.* **95,** 832–841
10. Massillon, D., Barzilai, N., Chen, W., Hu, M., and Rossetti, L. (1996) *J. Biol. Chem.* **271,** 9871–9874
11. Mithieux, G., Vidal, H., Zitoun, C., Bruni, N., Daniele, N., and Minassian, C. (1996) *Diabetes* **45,** 891–896
12. Granner, D. K., and O'Brien, R. M. (1992) *Diabetes Care* **15,** 369–395
13. O'Brien, R. M., and Granner, D. K. (1996) *Physiol. Rev.* **76,** 1109–1161
14. Massillon, D., Chen, W., Barzilai, N., Prus-Wertheimer, D., Hawkins, M., Liu, R., Taub, R., and Rossetti, L. (1998) *J. Biol. Chem.* **273,** 228–234
15. Argaud, D., Kirby, T. L., Newgard, C. B., and Lange, A. J. (1997) *J. Biol. Chem.* **272,** 12854–12861
16. Streeper, R. S., Svitek, C. A., Chapman, S., Greenbaum, L. E., Taub, R., and O'Brien, R. M. (1997) *J. Biol. Chem.* **272,** 11698–11701
17. Streeper, R. S., Eaton, E. M., Ebert, D. H., Chapman, S. C., Svitek, C. A., and O'Brien, R. M. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9208–9213
18. Li, Y., and van de Werve, G. (2000) *Biochem. Biophys. Res. Commun.* **272,** 41–44
19. Seoane, J., Trinh, K., O'Doherty, R. M., Gomez-Foix, A. M., Lange, A. J., Newgard, C. B., and Guinovart, J. J. (1997) *J. Biol. Chem.* **272,** 26972–26977
20. An, J., Li, Y., van De Werve, G., and Newgard, C. B. (2001) *J. Biol. Chem.* **276,** 10722–10729
21. Arden, S. D., Zahn, T., Steegers, S., Webb, S., Bergman, B., O'Brien, R. M., and Hutton, J. C. (1999) *Diabetes* **48,** 531–542
22. Perales, M. A., Sener, A., and Malaisse, W. J. (1991) *Mol. Cell. Biochem.* **101,** 67–71
23. Trandaburu, T. (1977) *Acta Histochem.* **59,** 246–253
24. Watanabe, A., and Nagashima, H. (1983) *Acta Med. Okayama* **37,** 463–470
25. Waddell, I. D., and Burchell, A. (1988) *Biochem. J.* **255,** 471–476
26. Khan, A., Chandramouli, V., Ostenson, C. G., Low, H., Landau, B. R., and Efendic, S. (1990) *Diabetes* **39,** 456–459
27. Khan, A., Hong-Lie, C., and Landau, B. R. (1995) *Endocrinology* **136,** 1934–1938
28. Ebert, D. H., Bischof, L. J., Streeper, R. S., Chapman, S. C., Svitek, C. A., Goldman, J. K., Mathews, C. E., Leiter, E. H., Hutton, J. C., and O'Brien, R. M. (1999) *Diabetes* **48,** 543–551
29. Sambrook, J., Fritsch, E. F., and Maniatis, E. F. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory, Plainview, NY
30. Jurka, J., Klonowski, P., Dagman, V., and Pelton, P. (1996) *Comput. Chem.* **20,** 119–121
31. Lacy, P. E., and Kostianovsky, M. (1967) *Diabetes* **16,** 35–39
32. Guest, P. C., Pipeleers, D., Rossier, J., Rhodes, C. J., and Hutton, J. C. (1989) *Biochem. J.* **264,** 503-v8
33. Kozak, M. (1991) *J. Biol. Chem.* **266,** 19867–19870
34. Bressan, G. M., and Stanley, K. K. (1987) *Nucleic Acids Res.* **15,** 10056
35. Jacoby, D. B., Zilz, N. D., and Towle, H. C. (1989) *J. Biol. Chem.* **264,** 17623–17626
36. Bischof, L. J., Martin, C. C., Svitek, C. A., Stadelmaier, B. T., Hornbuckle, L. A., Goldman, J. K., Oeser, J. K., Hutton, J. C., and O'Brien, R. M. (2001) *Diabetes* **50,** 502–514
37. Arden, S. D., Roep, B. O., Neophytou, P. I., Usac, E. F., Duinkerken, G., de Vries, R. R., and Hutton, J. C. (1996) *J. Clin. Invest.* **97,** 551–561
38. Jackson, I. J. (1991) *Nucleic Acids Res.* **19,** 3795–3798
39. Kollmar, R., and Farnham, P. J. (1993) *Proc. Soc. Exp. Biol. Med.* **203,** 127–139
40. Milpetz, F., Argos, P., and Persson, B. (1995) *Trends Biochem. Sci.* **20,** 204–205
41. Stukey, J., and Carman, G. M. (1997) *Protein Sci.* **6,** 469–472
42. Hemrika, W., Renirie, R., Dekker, H. L., Barnett, P., and Wever, R. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 2145–2149

43. Wobbe, C. R., and Struhl, K. (1990) *Mol. Cell. Biol.* **10,** 3859–3867
44. Stein, R. (2001) in *Handbook of Physiology. Section 7: The Endocrine System* (Jefferson, L. S., and Cherrington, A. D., eds) Vol. II, pp. 25–47, Oxford University Press, New York
45. German, M. (2000) in *Diabetes Mellitus: A Fundamental and Clinical Text* (LeRoith, D., Taylor, S. I., and Olefsky, J. M., eds) 2nd Ed., pp. 11–19, Lippincott-Raven Publishers, Philadelphia
46. Li, N., and Seetharam, B. (1998) *J. Biol. Chem.* **273,** 28170–28177
47. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) *Nucleic Acids Res.* **23,** 4878–4884
48. Winter, W. E. (2000) *Pediatr. Diabetes* **1,** 88–117
49. Habener, J. F., and Stoffers, D. A. (1998) *Proc. Assoc. Am. Physicians* **110,** 12–21
50. Sander, M., and German, M. S. (1997) *J. Mol. Med.* **75,** 327–340
51. Edlund, H. (1998) *Diabetes* **47,** 1817–1823
52. Bramblett, D. E., Huang, H. P., and Tsai, M. J. (2000) *Adv. Pharmacol.* **47,** 255–315
53. Pan, C. J., Lei, K. J., Annabi, B., Hemrika, W., and Chou, J. Y. (1998) *J. Biol. Chem.* **273,** 6144–6148

Cloning and Characterization of the Human and Rat Islet-specific Glucose-6-phosphatase Catalytic Subunit-related Protein (IGRP) Genes

Cyrus C. Martin, Larry J. Bischof, Barbara Bergman, Lauri A. Hornbuckle, Carl Hilliker, Claudia Frigeri, David Wahl, Christina A. Svitek, Randall Wong, Joshua K. Goldman, James K. Oeser, Frédéric Leprêtre, Philippe Froguel, Richard M. O'Brien and John C. Hutton

J. Biol. Chem. 2001, 276:25197-25207.
doi: 10.1074/jbc.M101549200 originally published online April 10, 2001
