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.M101849200

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 Lepreâtre**, 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

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL 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), AF324433 (rat IGRP promoter and exon 1), AF324434-AF324436 (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).

†† To whom correspondence may be addressed: Dept. of Molecular Physiology and Biophysics, 781 MB II, Vanderbilt University Medical School, Nashville, TN 37232-9615. 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.).

* 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 Nor-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.

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 micromeres, 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 G6Pase 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 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.

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 4q28–32 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...
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 glycoinosmetabolism.

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 db/db 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 b 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 expression between rodent species.

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—**[32P]-labeled 1028-bp cDNA probe isolated as a PsI 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 µl of 6X SSC (0.15 M NaCl and 0.015 M sodium citrate, 0.5% SDS, 100 µg/ml salmon sperm DNA, and 10 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 phage using cesium chloride gradient centrifugation (29). Two independent clones designated PAC 294 (~90 kb) and PAC 299 (~110 kb) 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 salI-PvuII fragment of a mouse IGRP cDNA clone (pSV.SPORT 1B1) containing exons 1, 2, and part of exon 3 (21) and an XhoI-XhoI 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 XhoI-XhoI fragment containing the promoter and exons 1–4 and an ~6-kbp EcoRI-EcoRI fragment containing exons 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 kb 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 (AC900045) containing a series of nonconiguous 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′-ATGGTGTGAGACCTAAGCCTAAG-3′ and reverse 5′-TGAAGTTTACATTCCTCTCCTC-3′) and exon 5 (forward 5′-AGACCGGATTTTCTATCTGCTC-3′ and reverse 5′-GGCGGCTATTTGCTCTGTG-3′) were used to amplify 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 mmt MgCl2 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).

**Localization of Rat IGRP Genomic Clones—**A fragment of the rat IGRP gene was generated using rat genomic DNA (CLONTENSE) as the template in a PCR reaction with the following primers: forward (5′-CGAATTCCCTCACAGATGTCGACCATACATG-3′) and reverse (5′-CGGATTACGCTGCAGTCCACAAATTG-3′); EcoRI 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 EcoRI 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 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). 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′-GGAATTACGCTACAGATGTCGACCATACATG-3′) and reverse (5′-CCCAAGATTCCCTCACAGATGTCGACCATACATG-3′).
Human and Rat IGRP

GCGTGAGGGGTTTGGAACTCACCTGCCAGC-3'); the EcoRI- and HindIII-cloning sites are underlined. These primers represent IGRF 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 EcoRI-HindIII-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 fragment was isolated within two overlapping genomic sub-clones; a 6-kbp Kpn1-Kpn1 fragment contained the promoter and exons 1–3, whereas a 6-kbp PstI-FsiI fragment contained exons 4 and 5. The identification of the exon/intron boundaries (Table I) was determined by direct DNA sequencing of both DNA strands 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 size of the PCR products was as expected. PCR reactions (100 μl) contained 100 pmol of each primer, 200 μM dNTP, 1.5 mM MgCl2, 200 ng of pGEM7 plasmid DNA containing the 6-kbp Kpn1-Kpn1 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 run for 94 °C for 5 min and then for 30 cycles of 30 s at 94 °C, 30 s at 53 °C, and 2 min at 72 °C followed by a final 20 min extension at 72 °C.

AmpliTaq DNA polymerase (PerkinElmer Life Sciences). Products were run for 94 °C for 5 min and then for 30 cycles of 30 s at 94 °C, 30 s at 53 °C, and 2 min at 72 °C followed by a final 20 min extension at 72 °C.

PCR reactions were performed with a PfuTurbo 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 PCR reaction was amplified using specific 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′-TCAAGATGGATTCCACAGAGGA-3′) and a compatible reverse primer located just 3′ to the stop codon (5′-CACAGGGTTACGGAGGGC-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′-GTAAGATGGATTAGAAGG-3′). The PCR products were inserted into the pTOPO blunt vector and then subcloned into pCDNA3.1 for expression studies using EcoRI and XhoI sites common to both vectors.

Generation of Antisera and Immunoperoxidase Staining—A PstI 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 is a useful fusion protein in reporter systems. Purified antibodies were raised in New Zealand white rabbits by immunization in incomplete Freund’s adjuvant followed by boosting in incomplete Freund’s adjuvant at six weekly intervals. Balb mouse pancreas was perfusion-fixed with 4%/w/v paraformaldehyde and subjected to standard paraffin embedding and sectioning before immunoperoxidase staining using the primary anti-semun diluted 1:100 in PBS (2 h at room temperature) and a secondary donkey anti-rabbit antibody conjugated to horseradish peroxidase (Jackson Laboratories).

Fusion Protein Masson 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 HindIII-PstI fragment and subcloned into HindIII-PstI-digested pSP72 (Promega). The resulting plasmid contains rat IGRP promoter sequence from −3, relative to the position of the transcription initiation site. The PstI site used in this cloning is conserved in the mouse IGRP (Fig. 2), 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 genes. A truncated rat IGRP-CAT fusion gene then generated by the enzyme digestion of the −900 IGRP-PstI fragment with HindIII and Nhel followed by Klenow treatment of the noncomplements ends and blunt-end ligation. The resulting plasmid has a calculated 5′ end point of −321. The TGTCA 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 primers as sense start points: the PstI site used for cloning purposes and the mutated base are underlined. The 5′ PstI primer (5′-CCGGATCCAGCTCTAGGAAAGC-3′), with the BamHI 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 HindIII-Nhel junction is shown in italics. The PCR fragment was digested with BamHI and PstI and subcloned into BamHI-PstI-digested pSP72 for sequencing. The promoter fragment was then re-isolated from the pSP72 plasmid as a BamHI-PstI fragment and ligated into BamHI-PstI-digested −321 rat IGRP-CAT. This BamHI site is located immediately 5′ of the HindIII cloning site in the pCAT(An) vector (35). A human IGRP-CAT fusion gene was generated by subcloning 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′-AACTGAGCGGTCCAGTTCGTTGTTCTTATGGGTCCTCCCCTGGTTGATC-3′) primers. HindIII and PstI sites used for cloning purposes are underlined. A single base pair change (italics) at position −1 in the human IGRP was introduced
into the 3′ primer to restore the PstI 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 HindIII and PstI and subcloned into HindIII-PstI-digested pSP72. The promoter fragment was then isolated from the pSP72 plasmid as a HindIII-BglII fragment and ligated into HindIII-BgII-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 with 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 PstI 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 59–1938) generated using the primers 5′-CAAGATGGAATTTCTCTGCTAGGATG-3′ (forward) and 5′-CAGAGCCTAATCTAGGCCACC-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-enzyme 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 MgSO4 (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 hydrolitic 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 ~1010 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).
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).

<table>
<thead>
<tr>
<th>Intron</th>
<th>Gene</th>
<th>5′ Intron junction</th>
<th>3′ Intron junction</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Mouse IGRP</td>
<td>TTTAAAT/Gtaagacct</td>
<td>acttacag/GATATTGT</td>
</tr>
<tr>
<td></td>
<td>Human IGRP</td>
<td>TTTAAAT/Gtaagatt</td>
<td>aatatacag/GATATTAT</td>
</tr>
<tr>
<td></td>
<td>Rat IGRP</td>
<td>TTTAAAT/Gtaagacct</td>
<td>gtttacag/GATATTGT</td>
</tr>
<tr>
<td>B</td>
<td>Mouse IGRP</td>
<td>AGGCCCAAG/gtaagcag</td>
<td>catctcag/GAACATCA</td>
</tr>
<tr>
<td></td>
<td>Human IGRP</td>
<td>AGTTTCCAG/gtaagcta</td>
<td>tgttcacag/GAACATCA</td>
</tr>
<tr>
<td></td>
<td>Rat IGRP</td>
<td>AGTTTCCAG/gtaagca</td>
<td>caatcag/GAACATCA</td>
</tr>
<tr>
<td>C</td>
<td>Mouse IGRP</td>
<td>CTGCACAC/gtacagctt</td>
<td>catcagac/AGCAGCTT</td>
</tr>
<tr>
<td></td>
<td>Human IGRP</td>
<td>CTGCACAC/gtacagctt</td>
<td>catcagac/AGCAGCTT</td>
</tr>
<tr>
<td></td>
<td>Rat IGRP</td>
<td>CTGCACAC/gtacagctt</td>
<td>Exon 4 absent</td>
</tr>
<tr>
<td>D</td>
<td>Mouse IGRP</td>
<td>GAATGTTT/Gtaaatat</td>
<td>atccacag/GGATGCTA</td>
</tr>
<tr>
<td></td>
<td>Human IGRP</td>
<td>AAATGTTT/Gtaaatat</td>
<td>atccacag/GGATGCTA</td>
</tr>
<tr>
<td></td>
<td>Rat IGRP</td>
<td>Exon 4 absent</td>
<td>atccacag/GGATGCTA</td>
</tr>
<tr>
<td></td>
<td>Consensus sequence:</td>
<td>(A or T)/G/taaa</td>
<td>cag/G</td>
</tr>
</tbody>
</table>

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 genes 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 AFM211y6) to D2S376 (microsatellite AFM319 × g1) (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 cDNA 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) and pI (8.62) of mouse IGRP and its size, determined pI 8.72.

Electrophoretic mobility of IGRP relates to the hydrophobic nature of the protein. The slower mobility of IGRP 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-
were translated in vitro T7 polymerase-derived transcripts from cloned pCDNA3.1 constructs.

Arg168; aa 179–193, Glu 191, aa 210–230, Arg 227; aa 255–280, amino acid(s) (aa): the sequence aa 25–47, Asp34 and Arg36; (TMAP (40)). Each stretch, however, contained a charged arginine; argues for conservation of function of IGRP. The G6Pase catalytic subunit 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, 3, 5, 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 Met57 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. 4A). Testis produced a weaker signal (10% of the pancreas signal) from hybridizing species of ~2400 and 1000 bp (Fig. 4B), 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. 4C) 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. 4D). 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. 4D) 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 $1/2$. 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.
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.

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).

Enzymatic 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 microsomal 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 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 TGTG 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. 7A 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
mouse promoter sequence located between −306 and +3. By contrast, neither the equivalent rat IGRP promoter sequence located between −321 and +3 (Fig. 7B) nor a longer fragment of the promoter containing sequence located between −900 and +3 (Fig. 7A) confers appreciable fusion gene expression. How-

ever, mutating the rat TGTA motif back to the consensus TATA motif markedly enhances basal rat IGRP-CAT fusion gene ex-

pression (Fig. 7B). Nevertheless, the actual level of rat pro-
moter activity remains 3-fold lower than that of the mouse (Fig. 7B), 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.

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 TC-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 (β-gal; 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.

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 IGRP 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 phys-

iologival 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 facili-
tate the identification of cis-acting elements, which are important for basal and islet-specific IGRP gene expression. Of spec-

ific 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. 4D). 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. 7B). 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 (41, 42). Yet as previously observed, no catalytic activity was obtained with either G6P or a generic phosphatase substrate (41, 42). Yet as previously observed, no catalytic activity was obtained with either G6P or a generic phosphatase substrate (41, 42).

**REFERENCES**

15. G6Pase construct, Kevin Docherty for NesY2 (human nesidioblastoma-derived insulin-secreting cell line) RNA, and Donna Curtis for assistance with the data base searching.
16. Human islet and rat IGRP was turned off because the protein was unnecessary, redundant, or deleterious.

**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 database searching. Human islets were obtained from Miami, FL, Rochester, MN and Boston, MA through the Juvenile Diabetes Foundation International islet consortium.

**Human and Rat IGRP**

3 B. Bergman and J. C. Hutton, unpublished observations.
Cloning and Characterization of the Human and Rat Islet-specific Glucose-6-phosphatase Catalytic Subunit-related Protein (IGRP) Genes

doi: 10.1074/jbc.M101549200 originally published online April 10, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M101549200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 51 references, 27 of which can be accessed free at http://www.jbc.org/content/276/27/25197.full.html#ref-list-1