Identification and characterization of a cDNA and the gene encoding the mouse ubiquitously expressed glucose-6-phosphatase catalytic subunit-related protein

J N Boustead, C C Martin, J K Oeser, C A Svitek, S I Hunter¹, J C Hutton¹ and R M O’Brien

Department of Molecular Physiology and Biophysics, 761 PRB, Vanderbilt University Medical School, Nashville, Tennessee 37232-0615, USA

¹Barbara Davis Center for Childhood Diabetes, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

(Requests for offprints should be addressed to R O’Brien; Email: richard.obrien@mcmail.vanderbilt.edu or J C Hutton; Email: john.hutton@uchsc.edu)

Abstract

Glucose-6-phosphatase (G6Pase) catalyzes the final step in the gluconeogenic and glycogenolytic pathways, the hydrolysis of glucose-6-phosphate (G6P) to glucose and phosphate. This paper describes the identification and characterization of a cDNA and the gene encoding the mouse ubiquitously expressed G6Pase catalytic subunit-related protein (UGRP). The open reading frame of this UGRP cDNA encodes a protein (367 amino acids (aa); M_r 38755) that shares 36% overall identity (56% similarity) with the mouse G6Pase catalytic subunit (357 aa; M_r 40454). UGRP exhibits a similar predicted transmembrane topology and conservation of many of the catalytically important residues with the G6Pase catalytic subunit; however, unlike the G6Pase catalytic subunit, UGRP does not catalyze G6P hydrolysis and does not contain a carboxy-terminal di-lysine endoplasmic reticulum retention signal. UGRP mRNA was detected by RNA blot analysis in every mouse tissue examined with the highest expression in heart, brain, testis and kidney. Database analysis showed that the mouse UGRP gene is composed of six exons, spans ~4.2 kbp of genomic DNA and is located on chromosome 11 along with the G6Pase catalytic subunit gene. The UGRP gene transcription start sites were mapped by primer extension analysis, and the activity of the mouse UGRP gene promoter was analyzed using luciferase fusion gene constructs. In contrast to the G6Pase catalytic subunit gene promoter, the UGRP promoter was highly active in all cell lines examined.

Journal of Molecular Endocrinology (2004) 32, 33–53

Introduction

Glucose-6-phosphatase (G6Pase) catalyses the hydrolysis of glucose-6-phosphate (G6P) to glucose and phosphate, the terminal step of the gluconeogenic and glycogenolytic pathways in liver. G6Pase is located in the endoplasmic reticulum (ER) as a multi-component system, but the exact number of components and their stoichiometry are unclear (van de Werve et al. 2000, Van Schaftingen & Gerin 2002). To date, two components have been identified - the catalytic subunit (Lei et al. 1993) and a G6P transporter (Gerin et al. 1997). Both components are integral membrane proteins that span the ER membrane multiple times. Two alternative models have been proposed for the structure of the G6Pase system. The first model, based mainly on rapid kinetic data, places the catalytic site within the membrane and ascribes both a transport function and catalytic activity to the catalytic subunit (van de Werve et al. 2000, Van Schaftingen & Gerin 2002). However, the validity of this rapid kinetic data has been challenged (Arion et al. 1998) and recent data favors an alternative substrate-transport model of G6Pase (Gerin et al. 2001). In this second model the G6Pase catalytic subunit has its catalytic site oriented towards the lumen of the ER and the G6P transporter serves to deliver G6P from the cytosol to the active site of the catalytic subunit (van de
Werve et al. 2000, Van Schaftingen & Gerin 2002). This latter model postulates the existence of transporters for inorganic phosphate and glucose that return the reaction products back to the cytosol; however, neither of these components have been identified.

Changes in hepatic G6Pase activity have a major impact on hepatic glucose production (HGP). Increased hepatic G6Pase activity is thought to contribute to the augmented HGP and the hyperglycemia that characterizes both type 1 and type 2 diabetes (Consoli 1992, Cline et al. 1994). Accordingly, G6Pase is overexpressed in animal models of diabetes (Liu et al. 1994, Haber et al. 1995). However, it is difficult to assess the relative contribution of increased G6Pase activity to the increased rate of HGP since the expression of many other catabolic enzymes also change in these diabetic animals (O’Brien & Granner 1996). Nevertheless, overexpression of the G6Pase catalytic subunit or the G6P transporter in hepatocytes (Seoane et al. 1997, Haber et al. 1995) or in vivo (Trinh et al. 1998) enhances the rate of G6P hydrolysis suggesting that increased G6Pase activity could make a significant contribution to the increased rate of HGP in diabetes. Conversely, decreased G6Pase activity caused by inactivating mutations within the G6Pase catalytic subunit and G6P transporter cause glycogen storage disease (GSD) types 1a and 1b respectively (Chou & Mansfield 1999), syndromes characterized by fasting hypoglycemia, hepatic glycogen accumulation and hepatomegaly.

Although G6Pase activity is highest in liver, it is also detected in kidney, intestine and islets (van de Werve et al. 2000, Van Schaftingen & Gerin 2002). The contribution of the kidney and intestine to the overall rate of gluconeogenesis is controversial (Stumvoll et al. 1997, Cherrington 1999, Mithieux 2001) and the role of G6Pase in islets is unclear (van de Werve et al. 2000, Van Schaftingen & Gerin 2002). Moreover, the G6Pase activity detected in islets displays distinct kinetic behavior and inhibitor profiles compared with that assayed in hepatic extracts (Arden et al. 1999), an observation that might be explained by the existence of a distinct G6Pase catalytic subunit isoform in islets, present in addition to the liver isoform (Arden et al. 1999). Indeed, we have previously identified a novel mouse cDNA (Arden et al. 1999) and gene (Ebert et al. 1999) that encodes an islet-specific G6Pase catalytic subunit-related protein (IGRP) which is similar in size (355 amino acids (aa); M_r 40 552), topology and sequence (~50% identity at the amino acid level) to the mouse G6Pase catalytic subunit. However, overexpression of IGRP in tissue culture cells does not increase G6P hydrolysis (Arden et al. 1999, Martin et al. 2001) and it is unclear whether IGRP catalyzes a distinct biochemical reaction or whether a cofactor or auxiliary protein is missing in the assay. In the course of searching for additional G6Pase catalytic subunit isoforms we identified a human cDNA and gene encoding a ubiquitously expressed G6Pase catalytic subunit-related protein (UGRP), a third member of the G6Pase gene family (Martin et al. 2002). Human UGRP is similar in size, predicted topology and sequence to the human G6Pase catalytic subunit but, as with IGRP, overexpression of UGRP in tissue culture cell studies does not increase G6P hydrolysis (Martin et al. 2002). Nevertheless, the exon/intron structures of all three genes are similar suggesting that they are evolutionarily related. In this paper we report the identification and characterization of a cDNA and the gene encoding mouse UGRP. The principal objectives of this study were to identify conserved amino acids that are potentially critical for UGRP function and conserved sequences within the UGRP gene promoter that may specify its ubiquitous tissue distribution.

Materials and methods

Materials

[α-32P]dATP (>3000 Ci/mmol) and [γ-32P]ATP (>5000 Ci/mmol) were obtained from Amersham. All individually specified reagents were of molecular biology grade and purchased from Sigma Chemical Company (St Louis, MO, USA). Mouse genomic DNA was purchased from Clontech (Palo Alto, CA, USA).

General cloning, DNA isolation and sequencing procedures

Plasmid DNA isolation, subcloning and restriction endonuclease analyses were performed by standard protocols (Sambrook et al. 1989). DNA fragments used for subcloning and labeling were isolated from...
agarose gels using either the QIAquick gel extraction kit (Qiagen) or Quantum Prep Freeze N Squeeze spin columns (Bio-Rad). DNA sequencing was performed using the USB (USB Corp, Cleveland, OH, USA) Sequenase kit or by automated sequencing using an ABI 310 DNA analyzer. Because transcription of the UGRP gene initiates at multiple start sites (see Fig. 6), all UGRP DNA sequences are numbered relative to the predicted translation start site. Plasmid constructs were purified either by centrifugation through cesium chloride gradients (Sambrook et al. 1989) (transcription analyses) or using commercial endotoxin-free plasmid isolation kits (Qiagen).

Endogenous UGRP mRNA expression analysis

The tissue distribution of UGRP mRNA was examined using commercially available blots containing mouse total RNA (see Fig. 3A; #MATB 1006–1, Seegene, Seoul, Korea), mouse polyA+ RNA (see Fig. 3B; #7762–1, Clontech), or human polyA+ RNA (see Fig. 3C,D; #7759–1 and #7755–1, Clontech). Mouse RNA blots were hybridized for 16 h at 65 °C in ExpressHyb hybridization solution (Clontech) with a 32P-radiolabeled randomly primed probe corresponding to the full length mouse UGRP open reading frame (ORF). This probe was isolated from the pcDNA3·1D/V5-His-TOPO vector (described below) as a BamH I-NotI fragment and then 25 ng were labeled by random oligonucleotide priming with [α-32P]dATP using the Stratagene (La Jolla, CA, USA) Prime-It II random primer labeling kit according to the manufacturer’s instructions. A 24 bp primer (5′-GATGCCCGCGCTGAGCGTGACTC-3′), complementary to mouse UGRP exon 1 sequence from +3 to +26, was synthesized by the Vanderbilt University Medical Center Diabetes Core laboratory and, following gel purification (Sambrook et al. 1989), was 5′-end-labeled with [γ-32P]ATP to a specific activity of ~2 Ci/µmol (Sambrook et al. 1989). The labeled primer (~3 × 105 c.p.m.) was annealed, for 1 h at 60 °C, to 50 µg total RNA isolated from various mouse tissues, as described above, and then primer extension was performed as previously described (Ebert et al. 1999).

Cell culture

Pancreatic islet β cell-derived βTC-3 and Min 6 cells, liver-derived HepG2 cells, cervix-derived HeLa cells, muscle-derived L6 cells and kidney-derived COS-7 cells were passaged as subconfluent cultures in Dulbecco’s modified Eagle’s medium (DMEM). Pancreatic islet β cell-derived INS-1 cells were passaged as subconfluent cultures in RPMI-1640 medium. Cell cultures were supplemented with 100 U/ml penicillin and 100 µg/ml streptomycin and with either 2·5% (vol/vol) fetal bovine serum and 15% (vol/vol) horse serum (βTC-3 cells), 15% (vol/vol) fetal bovine serum (Min 6 cells), 2·5% (vol/vol) newborn bovine serum, 2·5% (vol/vol) fetal bovine serum and 5% (vol/vol) Nu serum IV (BD Biosciences, Bedford, MA, USA) (HepG2 cells), 10% (vol/vol) newborn bovine serum (HeLa cells) or 10% (vol/vol) fetal bovine serum (L6, COS-7 and INS-1 cells). For transfection analyses L6 myoblasts were grown to near confluency and were then induced to differentiate by changing the medium to DMEM supplemented with 2% horse serum and 1 nM insulin 48 h before transfection, and then to DMEM supplemented with 1000 nM insulin alone 24 h before transfection.

Fusion gene plasmid construction and analysis

The construction of luciferase fusion genes containing mouse G6Pase catalytic subunit and mouse IGRP promoter sequence from −231 to
+66 and −306 to +3, respectively, relative to the transcription start sites, in the pGL3-Mod vector has been described (Martin et al. 2002). Likewise, the construction of a luciferase fusion gene containing human UGRP promoter sequence from −474 to +1, relative to the translation start site, in the pGL3-Mod vector has also been described (Martin et al. 2002).

A fragment of the mouse UGRP promoter from −497 to +1, relative to the translation start site was generated in a PCR reaction using mouse genomic DNA as the template and the following primers: forward (5′-CCCAAGCCTGGTAGTGA
t GTGGGAAAACAGAG-3′) and reverse (5′-GA
GATCTGGTGAATCAGAGTCTCCTAGCCT
GGG-3′). HindIII and BglII cloning sites underlined. The PCR product generated was digested with HindIII and BglII, ligated into a HindIII and BglII digested pSP72 vector (Promega) and then sequenced using the USB Sequenase kit to ensure the absence of polymerase errors. The sequence of this fragment was identical to that reported in the hts database for accession number AL954730. The HindIII-BglII UGRP promoter fragment was subsequently subcloned into the HindIII and BglII digested pGL3-Mod vector.

For fusion gene analyses, HepG2 and HeLa cells (in 8.5 cm diameter dishes) were co-transfected as previously described using a calcium phosphate precipitate containing 15 µg of a firefly luciferase fusion gene construct and 0.1 µg of a SV40-firefly luciferase fusion gene construct and 0.5 µg of a Renilla luciferase plasmid (Promega) as previously described (O’Brien et al. 1995, Chapman et al. 1999, Martin et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 2 µg of a firefly luciferase fusion gene construct and 0.5 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001). L6 cells (in 8.5 cm diameter dishes) were co-transfected with 10 µg of a firefly luciferase fusion gene construct and 0.1 µg of the SV40-Renilla luciferase plasmid using poly(l)-ornithine hydrobromide (Sigma) as previously described (Bischof et al. 2001).

UGRP protein expression analysis

The mouse UGRP ORF was amplified by RT-PCR from liver, kidney or βTC-3 insulinoma cell line total RNA preparations using the forward primer (5′-CACCAATGGACTCCAGCTGGG-3′) and, as a reverse primer, either (5′-TCAAGAGGAGGCAATGGGAT-3′) or (5′-AGAGGAGCGAATGG-3′). The forward primer incorporates a modified Kozak sequence required for cloning into the pcDNA3·1D/V5-His-TOPO vector (Invitrogen directional expression kit); the reverse primers use either the native stop codon or allow COOH terminal read through into the V5 epitope and (His)6 residues of the vector. DNA sequencing showed that the mouse cDNAs prepared from different sources were identical and that the inserted ORF was identical to mouse Unigene Mus musculus 22385. The mouse G6Pase catalytic subunit, mouse IGRP and β-galactosidase cloned into the same vector served as control constructs (Arden et al. 1999, Martin et al. 2001).

In vitro transcription/translation assays were performed using rabbit reticulocyte lysate with the TNT Quick coupled transcription/translation kit (Promega) as previously described (Arden et al. 1996) using T7 polymerase transcripts generated from the cloned mouse G6Pase catalytic subunit, mouse IGRP and mouse UGRP sequences. The 35S-methionine labeled products were separated on 12.5% SDS-polyacrylamide gels and visualized by direct phosphorimaging (Molecular Dynamics STORM phosphorimager).

The mouse G6Pase catalytic subunit, mouse IGRP and mouse UGRP proteins were expressed by transient transfection of COS-7 cells using a commercial lipofection reagent (Fugene; Roche). The pcDNA3·1 plasmids (2 µg) together with 1 µg β-galactosidase plasmid were mixed with 9 µl Fugene 6 reagent diluted in 100 µl serum-free DMEM media and allowed to stand at room temperature for 15 min before addition to 6 × 105 COS-7 cells that had been freshly plated in 4 ml DMEM containing 10% (v/v) fetal bovine serum and antibiotics in 6-cm Petri dishes. Cells were harvested after 48 h using a non-enzymic
procedure (GIBCO cell dissociation buffer, Life Technologies), rinsed twice in 150 mM NaCl, 10 mM Heps, 1 mM EDTA (pH 6·8) then resuspended in 1 ml 0·27 M sucrose, 10 mM 2-[N-Morpholino]ethanesulphonic acid, potassium salt (MES), 2 mM EGTA, 1 mM MgSO4 (pH 6·5) and then sonicated for 20 s. The sonicate was centrifuged at 8000 g for 6 min in a refrigerated Eppendorf microfuge to remove debris and a particulate fraction prepared by further centrifugation of the supernatant at 50 000 r.p.m. for 30 min in a Beckman TLN-110 rotor. The pellet was resuspended in 200 µl homogenization media (approx. 0·2–0·5 mg/ml protein) and assayed for G6P hydrolytic activity in a 100 mM dimethylglutarate buffer (pH 6·5) containing 10 mM Na K tartrate, 10 mM EDTA and 10 mM G6P as previously described (Arden et al. 1999).

To evaluate the level of expression of each of the constructs in COS cells, plasmids encoding V5-His tagged variants of the mouse G6Pase catalytic subunit, mouse IGRP and mouse UGRP were transiently transfected as described above. Cell homogenates (100 µg protein) were electrophoresed on 12·5% SDS-polyacrylamide gels, transferred to nitrocellulose as previously described (Hutton et al. 1988) and subjected to immunoblot analysis using a horseradish peroxidase (HRP) conjugated anti-V5 antibody (Invitrogen) as previously described (Hutton et al. 1988).

To evaluate the level of expression of endogenous UGRP in tissues and cell lines a UGRP antiserum was generated. New Zealand white rabbits were immunized with a 14 amino acid peptide HTLSDQEAPPIRSS, representing the C-terminus of mouse UGRP, bearing an N-terminal cysteine for conjugation to keyhole limpet hemocyanin in Freunds complete adjuvant. Animals were boosted at 30 days and bled out at 75 days. Protein samples (25 µg protein) were electrophoresed on 12·5% SDS-polyacrylamide gels and after transfer to nitrocellulose (Hutton et al. 1988) blotted overnight with a 1:10 000 dilution of primary antibody in TBS (10 mM Tris, pH 7·6, 150 mM NaCl), 0·1% Tween-20, 3% milk powder. Primary antibody binding was then detected using the ECL Plus Western Blotting kit (Amersham Life Science; # RPN2124) using a 1:5000 dilution of a secondary HRP conjugated anti-rabbit antibody.

**Immunohistochemical staining**

Pancreatic tissue was fixed in 4% paraformaldehyde in PBS and then embedded in optimal cutting temperature compound (OCT) before frozen sectioning. Eight-micron sections were stained with 1:500 dilutions of the following primary antibodies: rabbit anti-UGRP, guinea pig anti-bovine insulin (Bright et al. 2002) and mouse monoclonal anti-glucagon (Abcam, Cambridge, Cambs, UK). The staining of all the islet cells and acinar cells with the UGRP antiserum was reduced to background by preabsorption of the antibody with 50 µg/ml of the peptide immunogen. Sections were washed in PBS 1% BSA and primary antibody binding was then detected using 1:500 dilutions of the following secondary antibodies: anti-guinea pig (labeled with Cy2), anti-rabbit (labeled with Cy3) and anti-mouse (labeled with Cy5) all obtained from The Jackson Laboratory (Bar Harbor, ME, USA). Images were recorded on a Nikon microphot FXA epifluorescence microscope with a monochrome digital camera (Roper Micromax) via Intelligent Imaging Systems software. The photos are displayed in pseudocolor.

**Results**

**Identification and sequence analysis of the mouse ubiquitously expressed G6Pase catalytic subunit-related protein (UGRP)**

A BLAST search of the NCBI nr database using the previously described human UGRP cDNA sequence (accession number BC002494; I.M.A.G.E. CloneID 3050476; Martin et al. 2002) as the query identified a cDNA (accession number AK002966) encoding mouse UGRP. Based on comparison with human UGRP, this 1408 bp cDNA would be predicted to encode the entire mouse UGRP ORF were it not for a single base pair deletion at position +362 relative to the predicted translation start site at +1. This deletion is almost certainly a sequencing/cloning error since a BLAST search of the NCBI mouse_est database using this mouse UGRP cDNA sequence identifies 18 mouse ests all of which contain an additional ‘C’ at base pair +362. In addition, analysis of the
91 bp putative 5′ untranslated leader sequence in the AK002966 cDNA identifies an in frame stop codon located 9 bp 5′ of the putative initiation methionine. This 9 bp sequence contains no other ATG codons. Moreover, the predicted UGRP initiation methionine is located in the context of a strong Kozak consensus sequence (Kozak 1991).

Since UGRP gene transcription initiates at multiple locations (see the primer extension analysis below) a unique full length mouse UGRP cDNA does not exist. Indeed, this conclusion is supported by an analysis of the NCBI mouse_est database that identifies multiple UGRP cDNAs with variable 5′ end points (see accession numbers BQ885400, BG914739, BG970809 for examples). This analysis also revealed that the mouse UGRP cDNA in the NCBI nr database (accession number AK002966) differs from all the cDNAs in the mouse_est database, based on the available sequence, in that it contains an additional exon that is generated by splicing of the 5′ untranslated region (see Fig. 7). The other cDNAs in the mouse_est database do not extend as far 5′ and instead they all terminate in what would be intron A of this longer mouse primary transcript. If the true UGRP transcription start was the one used to generate the UGRP mRNA represented by AK002966 then all these other cDNAs in the mouse_est database would represent short transcripts all of which were partially processed. The latter would be highly unusual since analysis of other gene products suggests that the mouse_est database typically contains fully processed but truncated transcripts. In addition, UGRP cDNAs containing this alternate exon were not found in the human_est database (Martin et al. 2002). Since use of this additional exon is a very rare event, similar to the rare alternative splicing of the 5′ untranslated region in the rat insulin II gene (Soares et al. 1986), it was not investigated further.

The deduced mouse UGRP ORF (Fig. 1) encodes a 346 amino acid protein (Mr 38 785, pI 8-53) of a generally hydrophobic character (226/346 residues). The hydrophobic amino acids are arranged in 9 major stretches, 8 of which are predicted to be able to span a phospholipid bilayer as an alpha helix (TMAP, Milpetz et al. 1995; sec: http://condor.bcm.tmc.edu; data not shown). Six of these stretches however, contain charged amino acid(s) (aa): the sequence from aa 25–49, Asp34 and Lys36; aa 53–77, Glu65 and Lys72; aa 144–166, Arg161; aa 247–273, Arg253 and Asp254; aa 278–297, Lys283; and residues 308–332, Lys316. The three putative transmembrane segments without charged residues are located between aa 115 and 135, 172 and 188, and 196 and 222. With the exception of Asp34, all the above mentioned charged amino acids are conserved in the human UGRP ORF which shows an overall 84% sequence identity (93% similarity). Mouse UGRP can be aligned with residues 5–357 of the mouse G6Pase catalytic subunit (36% identity; 55% similarity) and residues 1–355 of mouse IGRP (35% identity; 55% similarity) the gaps occurring in non-conserved regions between the latter molecules. IGRP and the G6Pase catalytic subunit are more closely related (50% identity; 75% similarity) (Fig. 1). The conservation of sequence particularly within hydrophobic segments and the charged residues within them suggest that they may have a function other than simple membrane spanning. UGRP contains no consensus sites for NH2-linked glycosylation (Jackson et al. 1990) unlike the G6Pase catalytic subunit (n=3), and IGRP (n=2), nor does it share the COOH-terminal consensus sequence (KKXX) found in G6Pase, IGRP and many ER-resident transmembrane proteins (Kornfeld & Kornfeld 1985). However, the absence of these motifs would not preclude localization of UGRP to the ER compartment (Jackson et al. 1990).

UGRP, like the G6Pase catalytic subunit and IGRP has an extended sequence motif found in bacterial vanadate-sensitive haloperoxidases and mammalian phosphatidic acid phosphatases (Hemrika et al. 1997, Stukey & Carman 1997) that constitutes the active site of these enzymes. The equivalent residues in mouse UGRP are Lys72=Arg Pro60=Pro Ser Gly His114=Ser Arg161=His170. Amino acids within the G6Pase catalytic subunit sequence whose mutation results in loss of function in GSD type 1a are shown in Fig. 1 (Chou & Mansfield 1999). Among the 41 aa residues in which inactivating mutations have been described in GSD type 1a, there are 14 positions that differ between mouse G6Pase catalytic subunit and mouse UGRP viz. T16=A12, K54>S30, G68=A64, T108>S103, G122>D117, A124=G119, G223=L212, T255>S245, N264>D254, L263=D255, R299>K283, V338=L325, I341>V328 and L345>V332. Human G6Pase and UGRP differ also at the same positions and, in addition, show differences at W156=L147, and S298=C286. By
contrast, mouse G6Pase and IGRP differ at only 6 of these positions.

**Translation and enzymic activity of mouse UGRP**

*In vitro* translation of a mouse UGRP cRNA generated a protein of 30 kDa (34 kDa with COOH terminal extension), a size somewhat smaller than the predicted $M_r$ (38 785 Da) (Fig. 2). The size discrepancy between the predicted and observed $M_r$ of mouse UGRP probably relates to the hydrophobic nature of the protein. The slower electrophoretic mobility of UGRP (30 kDa) compared with mouse IGRP (36 kDa) and the mouse G6Pase catalytic subunit (33 kDa) is conceivably related to a combination of the length of the ORF (346, 355 and 357 aa respectively) and the more acidic nature (predicted pI 8.53, 8.72 and 9.22 respectively).

**Figure 1** Alignment of the deduced peptide sequences of the mouse G6Pase catalytic subunit, IGRP and UGRP.

The alignment of the predicted amino acid sequence of mouse UGRP (accession number AK002966; Mm. 22 385), mouse IGRP (accession number Z47787; Mm. 140 768) and the mouse G6Pase catalytic subunit (accession number U01120; Mm. 18 064) is shown. Identities are indicated by solid circles and similarities by vertical bars. Residues defined as important to catalytic activity in haloperoxidases and related phosphatases are shown in blue. Inactivating point mutations in the human G6Pase catalytic subunit that abolish (black letters) or reduce (red letters) catalytic activity and give rise to GSD type 1a (Chou & Mansfield 1999) are indicated in single letter code above the UGRP sequence.
mouse UGRP ORFs linked to a human growth hormone 3′UTR and polyadenylation signal (pcDNA 3·1D v5-His-TOPO vector). Constructs were generated which either terminated at the native stop codon or encoded a protein, designated V5 His (Fig. 2, Table 1), with a COOH terminal extension provided by the vector that includes a 14 aa V5 epitope and six His residues followed by a stop codon. The rationale behind this approach was to inactivate the endoplasmic reticulum localization signal that is needed for targeting to the ER and which is conceivably important for expression of enzymic activity. The efficiency of transfection was evaluated by co-transfection of a CMV-β-galactosidase construct also in the same vector. In addition, immunoblot analyses were performed to demonstrate that

Table 1  G6P hydrolytic activity in microsomal fractions derived from transiently transfected COS-7 cells. COS-7 cells (approx. 5×10⁵/assay) were co-transfected, as described in Materials and methods with the indicated pcDNA3.1 plasmid encoding mouse (m) UGRP, mouse IGRP or the mouse G6Pase catalytic sub unit (2 µg) together with a reference vector encoding β-galactosidase (1 µg). Enzyme activities were determined in cell homogenates after 48 h as described in Materials and methods. Each tabulated result represents the mean±S.E.M. of duplicate determinations from the indicated number of experiments. V5 His constructs are COOH terminally extended versions of the molecules that incorporate a V5 epitope and His₆ tag generated by deletion of the stop codon in each insert.
the V5 His variants of the mouse G6Pase catalytic subunit, mouse IGRP and mouse UGRP were all expressed at similar levels in COS cells (Fig. 2B).

Transfection with the G6Pase catalytic subunit construct resulted in a ~100-fold increase in G6P hydrolysis over basal activity (Table 1) without obvious deleterious effects on the growth or morphology of the cells. In contrast, transfection with constructs encoding mouse UGRP or IGRP did not produce a significant change in G6P hydrolytic rates (Table 1). The rates of G6P hydrolysis with UGRP and IGRP transfection tended to exceed the rates observed in mock-transfected and untransfected control cells; however, this was rarely more than twofold that of the control and did not reach statistical significance. The activity of transfected β-galactosidase in these cells was not altered by the co-transfection of the G6Pase catalytic subunit, IGRP or UGRP encoding plasmids, thus indicating that neither cell viability nor biosynthetic capacity was adversely affected by the introduction of these constructs (Table 1).

The rate of hydrolysis of a series of low molecular weight phosphorylated substrates (p-nitrophenol phosphate, mannose-6-phosphate, pyrophosphate, ribose-5-phosphate, 6-phosphogluconate, glucose-1-phosphate, glucose-1,6 diphasphate, fructose-1-phosphate, fructose-6-phosphate, erythrose-4-phosphate, glucosamine-6-phosphate, phosphatidic acid, Mg2+ ATP, Mg2+ ADP and Mg2+ AMP) many of which can be hydrolyzed by the G6Pase catalytic subunit (Arden et al. 1999) was not significantly increased by the introduction of UGRP (data not shown). Furthermore, enzyme assays performed after pretreatment of the enzyme with the purified catalytic subunit of protein kinase A and Ca2+/calmodulin-dependent kinase II did not reveal any hydrolytic activity associated with UGRP (data not shown). The presence of Ca2+ (5 mM) also failed to stimulate G6P hydrolysis by UGRP (data not shown). Finally, the addition of a COOH terminal V5 epitope/His tag that would likely affect the localization of the G6Pase catalytic subunit did not impede its hydrolytic activity significantly nor affect G6P hydrolysis in IGRP and UGRP transfected cells (Table 1).

**Tissue distribution and expression of mouse UGRP mRNA**

RNA blot analyses of total mRNA using the mouse UGRP ORF described above as the labeled probe showed the presence of a single ~1550 nucleotide hybridizing species in 13 major mouse tissues (Fig. 3A). Expression in some tissues was only apparent after longer exposure of the film (data not shown). The calculated size of mouse UGRP mRNA without a poly A+ tail is 1318 nucleotides (Table 2) a size that is consistent with the ~1550 nucleotide hybridizing species seen on RNA blots (Fig. 3A and B), given that the UGRP cDNAs reported in the NCBI mouse_est database contain variable 5’ untranslated leader sequences up to ~240 nucleotides in length. The highest level of mouse UGRP expression was observed in heart, brain, kidney and testis and intermediate levels were detected in lung, spleen, stomach, small intestine, skeletal muscle and uterus (Fig. 3A). The lowest levels of UGRP expression were found in liver, thymus and placenta; expression in these tissues was only clearly apparent after longer exposure of the film (data not shown). However, the RNA blot shown in Fig. 3B, using polyA+ RNA, confirms that UGRP is expressed in liver, but at a lower level than in other tissues. The most notable expression of UGRP seen in Fig. 3A is in testis. However, on the RNA blot shown in Fig. 3B UGRP expression in testis is similar to that in heart and brain whereas in Fig. 3A it is greater in testis than in these two tissues. Although speculative, this variability may indicate that UGRP gene expression is subject to hormonal/nutrient regulation. Figure 3C shows that UGRP is also expressed in human testis, prostate, ovary and colon, additional human tissues that had not been examined in our previous characterization of the human UGRP gene (Martin et al. 2002). Figure 3D shows that UGRP is expressed in all brain regions examined. These results demonstrate that UGRP is ubiquitously expressed. This broad tissue distribution of UGRP contrasts with that of the G6Pase catalytic subunit and IGRP whose distributions are highly restricted (van de Werve et al. 2000, Martin et al. 2001, Van Schaftingen & Gerin 2002).

**Tissue distribution and expression of mouse UGRP protein**

Although UGRP mRNA has been detected in all tissues so far examined, the question as to whether UGRP is equally expressed in all cell types within those tissues remains to be addressed. To begin to
assess the distribution of UGRP within different tissues, an antiserum was raised to a peptide representing the C-terminal 14 amino acids of mouse UGRP. Figure 4 (upper panel) shows that, when kidney-derived COS or pancreatic islet cell-derived INS-1 cells are transfected with a plasmid encoding native mouse UGRP, an increased abundance of a ~30 kDa protein is detected by the UGRP antiserum. The molecular mass of this protein matches that of native UGRP as assessed by in vitro transcription and translation (Fig. 2A). This ~30 kDa protein is also detected in mouse pancreatic islets and pancreatic islet β cell-derived Min 6 cells (Fig. 4, upper panel), as well as in brain, thymus, lung, kidney, spleen and pancreas (Fig. 4, lower panel). The relative abundance of this ~30 kDa protein in brain relative to these other tissues is consistent with the higher abundance of UGRP mRNA in brain (Fig. 3). The UGRP antiserum also detects a faster migrating protein of ~26 kDa (Fig. 4). However, the abundance of this species varies between different preparations of extracts (compare the Min 6 lanes in Fig. 4 upper and lower panels),
suggested that it may represent a proteolyzed form of UGRP. 

Since UGRP protein was detected in islets (Fig. 4) we next used the UGRP antiserum to compare its relative abundance in pancreatic endocrine and exocrine tissue. Immunohistochemistry showed bright UGRP staining of islets including overlap with alpha and beta cells (Fig. 5). UGRP staining was also detected in insulin- and glucagon-negative cells suggesting that UGRP is also expressed in other islet cell types (Fig. 5). This pattern of UGRP expression contrasts with that of IGRP which is expressed predominantly in islet beta cells (Hutton & Eisenbarth 2003). Figure 5 also shows a significant UGRP signal coming from the exocrine tissue suggesting that these cells also contain UGRP. This conclusion is consistent with the results of the Western blots. Thus, although a UGRP signal was obtained from both pancreas and isolated pancreatic islets (Fig. 4), it is evident that the pancreas sample must contain an extra-islet source of UGRP because islets comprise only 1–2% of the pancreatic cell mass.

**Identification and structural analysis of the mouse UGRP gene**

A BLAST search of the NCBI htgs database using the mouse UGRP cDNA sequence (accession number AK002966) as the query identified a BAC clone (accession number AL954730; approximately 193 kbp) which contains the entire mouse UGRP gene.

### Table 2 Structure of the mouse G6Pase catalytic subunit, IGRP and UGRP genes. The mouse UGRP gene exon and intron sizes as well as the G6Pase catalytic subunit and IGRP intron sizes were determined as described in Results. The lengths of the coding sequences in exon 1 of the mouse and human G6Pase catalytic subunit, IGRP and UGRP genes are 230 bp, 218 bp and 219 bp respectively. Similarly, the lengths of the coding sequences in exon 5 of the mouse and human G6Pase catalytic subunit and IGRP genes are both 509 bp whereas the lengths of the coding sequences in exon 6 of the mouse and human UGRP genes are both 361 bp.

<table>
<thead>
<tr>
<th>Exon number</th>
<th>Exon size (bp)</th>
<th>Intron size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G6Pase</td>
<td>IGRP</td>
</tr>
<tr>
<td>1</td>
<td>311</td>
<td>254</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>3</td>
<td>106</td>
<td>112</td>
</tr>
<tr>
<td>4</td>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>5</td>
<td>1615</td>
<td>1284</td>
</tr>
<tr>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The mouse G6Pase catalytic subunit and mouse IGRP exon sizes are from Shelly et al. (1993) and Ebert et al. (1999), respectively. N/A, not applicable.

Figure 4 Analysis of mouse UGRP protein expression by Western blotting. UGRP expression in the indicated tissues and cell lines was analyzed by immunoblot analysis as described in Materials and methods. In addition, UGRP expression was compared in COS-7 and INS-1 cells that were mock transfected or transiently transfected with a plasmid encoding native mouse UGRP. Representative results are shown.
gene as a contiguous sequence. This comparison of UGRP cDNA and genomic sequences broadly mapped the exon/intron boundaries of the UGRP gene. The precise exon/intron splice site boundaries (Table 3) were then determined by comparison with the splice consensus sequence (Jackson 1991) and, where the sequence differed from the consensus, by the requirement that the UGRP open reading frame be maintained. The sequences of the six UGRP exons in the UGRP gene, as determined from genomic DNA, were identical to the equivalent sequences in the UGRP cDNA with the exception of the single base pair deletion in the cDNA that represents a sequencing/cloning error (see above) and a single base pair difference in exon 5. A BLAST search of the NCBI mouse_est database indicates that this latter difference also probably represents a second sequencing/cloning error in the cDNA.

A similar BLAST search of the NCBI nr database using the mouse G6Pase catalytic subunit cDNA sequence (accession number U00445; Shelly et al. 1993) as the query identified a BAC clone (accession number AL590969; approximately 250 kbp) which contains the entire mouse G6Pase catalytic subunit gene as a contiguous sequence. Thirteen differences were apparent between the sequences of the five G6Pase catalytic subunit exons as determined from genomic DNA and the equivalent sequences reported for the cDNA (Shelly et al. 1993). Inspection of multiple other entries in the NCBI nr and mouse_est databases suggests that some of these differences represent polymorphisms whereas others may be sequencing/cloning errors in the cDNA.

Once the UGRP and G6Pase catalytic subunit exon/intron boundaries were determined, the sizes of the individual exons and introns could then be calculated (Table 2). The sizes of the mouse G6Pase catalytic subunit gene exons have been described previously (Shelly et al. 1993) and were confirmed in this analysis. A summary of the sizes of the individual exons and introns is shown in Table 2. Analysis of the equivalent mouse UGRP genomic sequence in the Celera database gave very similar estimates of intron sizes whereas the mouse G6Pase catalytic subunit genomic sequence is currently not represented in the Celera database.

The mouse IGRP exon/intron boundaries and the sizes of the individual exons have been determined previously and the sizes of three of the four individual introns estimated using PCR (Ebert et al. 1999). To determine the intron sizes precisely a BLAST search of the NCBI nr database using the mouse IGRP cDNA sequence (accession number Z47787; Arden et al. 1999) as the query identified a BAC clone (accession number AC084429) which contains the entire mouse IGRP gene as a contiguous sequence. The sequences of the five IGRP exons, as determined from genomic DNA, were identical to the equivalent sequences reported in the cDNA (Arden et al. 1999) with the exception of a single base pair in the 3′ untranslated region. A BLAST search of the NCBI mouse_est database indicates that this difference probably represents a polymorphism. The sizes of the individual introns were close to those previously estimated by PCR (Table 2; Ebert et al. 1999).
Table 2 reveals a general conservation of exon sizes between all three genes, an observation that is consistent with the hypothesis that all three genes are evolutionarily related. The mouse UGRP gene, however, like the human UGRP gene (Martin et al. 2002), has an additional exon (Table 2) which appears to have arisen from insertion of an intron in the UGRP gene within the equivalent of exon 5 of the G6Pase catalytic subunit or IGRP gene (or alternatively its excision from an ancestral G6Pase catalytic subunit or IGRP gene). The other notable difference is that the 3′/5′ untranslated region in the mouse UGRP gene, like the human UGRP gene (Martin et al. 2002), is shorter than that of either the G6Pase catalytic subunit or IGRP genes. The 3′ untranslated regions of the mouse G6Pase catalytic subunit and IGRP mRNAs are 1106 and 775 respectively and are both encoded by exon 5. In contrast, the 3′ untranslated region of mouse UGRP mRNA is 280 and is encoded by exon 6. This calculation of the length of the mouse 3′ untranslated region is based on the analysis of the 1408 bp mouse UGRP cDNA described by accession number AK002966 which was cloned using a 3′ primer with a poly T tract. A consensus polyadenylation signal sequence AAUAAA (Colgan & Manley 1997) is located 20 bp 5′ of the end of this cDNA. In addition, multiple other UGRP cDNAs in the NCBI mouse_est database have 3′ untranslated regions of almost identical lengths.

Table 3 Comparison of the exon/intron boundaries of the mouse G6Pase catalytic subunit, IGRP and UGRP genes. The mouse IGRP exon/intron boundaries are from Ebert et al. (1999) whereas the mouse G6Pase catalytic subunit and UGRP exon/intron boundaries were determined as described in Results. Exon and intron sequences are shown in uppercase and lowercase letters respectively. The 5′ and 3′ consensus splice sequences are from Jackson (1991).

<table>
<thead>
<tr>
<th>Intron</th>
<th>Gene</th>
<th>3′ exon/5′ intron junction</th>
<th>3′ intron/5′ exon junction</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>G6Pase</td>
<td>T T C A A G T G/ g t a a g a a c</td>
<td>c t c c a c a g/G A T T C T G T</td>
</tr>
<tr>
<td></td>
<td>IGRP</td>
<td>T T A A A T G/ g t a a g a c t</td>
<td>a c t c a c a g/G A T A T G T</td>
</tr>
<tr>
<td></td>
<td>UGRP</td>
<td>T T C A A G T G/ g t g a g a a</td>
<td>t c t t c c a g/G T T T C T G T</td>
</tr>
<tr>
<td>B</td>
<td>G6Pase</td>
<td>C G G A C C A G/ g t g a g a g t</td>
<td>c c a t g t a g/G A A G T C C</td>
</tr>
<tr>
<td></td>
<td>IGRP</td>
<td>A G G C C A C A G/ g t a a g c a g</td>
<td>c a t t g c a g/G A A G T C C</td>
</tr>
<tr>
<td></td>
<td>UGRP</td>
<td>T G G T C C A G/ g t g g g a a a</td>
<td>c c c g g a c/G C A G C C C</td>
</tr>
<tr>
<td>C</td>
<td>G6Pase</td>
<td>G G A T T C C G/ g t a a g a c</td>
<td>t c t t g c a g/G T G T T T G A</td>
</tr>
<tr>
<td></td>
<td>IGRP</td>
<td>C T G C A C A G/ g t c a g c t t</td>
<td>c a t c a c a g/A C T G A C T</td>
</tr>
<tr>
<td></td>
<td>UGRP</td>
<td>T C C G C C A G/ g t a g a c c</td>
<td>c t t t c c a g/C C C C T G G G</td>
</tr>
<tr>
<td>D</td>
<td>G6Pase</td>
<td>C T T G T C A G/ g t a t c a g c</td>
<td>c c a c a c a g/G C A T T G C T</td>
</tr>
<tr>
<td></td>
<td>IGRP</td>
<td>G A T T G G T G/ g t a a a t t a t</td>
<td>a t c c c c a g/G G A T G C T A</td>
</tr>
<tr>
<td></td>
<td>UGRP</td>
<td>T A T T G T G/ g t a a g c a a</td>
<td>t c c t t a g/G T G C T G C C</td>
</tr>
<tr>
<td>E</td>
<td>G6Pase</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>IGRP</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>UGRP</td>
<td>C T T T C C T G/ g t a a g t c c</td>
<td>c c c c a c a g/G T C C A T C A</td>
</tr>
<tr>
<td>Consensus sequence</td>
<td>(A or T)G/ g t a a</td>
<td>c a g/G</td>
<td></td>
</tr>
</tbody>
</table>

N/A not applicable.
than the G6Pase catalytic subunit or IGRP genes (Table 2). Consequently, the mouse G6Pase catalytic subunit gene spans \(10.2\) kbp compared with \(7.9\) kbp for mouse IGRP and \(4.2\) kbp for mouse UGRP (Table 2). The sizes of the introns also vary between the mouse and human genes encoding the G6Pase catalytic subunit, IGRP and UGRP (compare Table 2 and Martin et al. 2002). Thus, the human G6Pase catalytic subunit gene spans \(12.1\) kbp compared with \(8.7\) kbp for human IGRP and \(5.4\) kbp for human UGRP (Table 2).

The exon/intron boundaries are generally well conserved between the mouse (Table 3) and human (Martin et al. 2002) G6Pase catalytic subunit, UGRP and IGRP genes and comply with consensus splice sequences (Jackson 1991) with the exception of the junction between intron C and the 5′ end of exon 4. All three genes have distinct boundary sequences here and only the G6Pase catalytic subunit gene has the consensus ‘G’ at this position (Table 3). Exon 4 is commonly spliced out in mouse (Arden et al. 1999) and human (Martin et al. 2001) IGRP which might relate to this difference. The spliced variants of IGRP (Arden et al. 1999) and that involving exon 2 of the G6Pase catalytic subunit (Haber et al. 1995) encode truncated proteins that do not contain the phosphatase consensus sequence (Fig. 1).

Chromosomal mapping of the mouse UGRP gene

We have previously mapped the location of the mouse IGRP gene to the proximal portion of mouse chromosome 2 near the marker D2 Mit11, positioned at 39 cM (Ebert et al. 1999). The chromosomal location of the gene encoding the mouse G6Pase catalytic subunit has not been reported. However, based upon the close homology between the human and mouse genomes and the observation that the human gene maps to chromosome 17 (Lei et al. 1994), the mouse ortholog for the G6Pase catalytic subunit would be expected to be on chromosome 11. Indeed, the BAC clone described above (accession number AL590969), that contains the entire mouse G6Pase catalytic subunit gene, represents mouse chromosome 11 sequence. Further analysis using the Celera Discovery System and Celera’s associated databases confirmed that the genes encoding mouse UGRP and mouse IGRP map to chromosomes 11 and 2 respectively.

Determination of the mouse UGRP gene transcription start site

To determine the transcription start site of the mouse UGRP gene, a primer extension analysis was performed using total RNA isolated from various mouse tissues (Fig. 6). The primer for this analysis was designed to be complementary to the mouse UGRP coding sequence between +3 and +26 (Fig. 7). Multiple transcription start sites were detected using this primer (Fig. 6). This is consistent with the existence of multiple UGRP cDNAs in the mouse_est database that contain variable 5′ untranslated leader sequences that map to the same region of the UGRP gene as detected by the primer extension analysis. Identical UGRP transcription start sites were detected in all mouse RNA samples assayed (Fig. 6) whereas the abundance of the primer extension reaction products varied in a tissue-specific manner consistent with the variation of UGRP gene expression detected by RNA blotting (Fig. 3). This observation serves as an important control in that it suggests that the UGRP primer extension primer is specifically recognizing UGRP mRNA.

The result of this primer extension assay suggests that the UGRP promoter is located immediately adjacent to the translation initiation codon in a region without a consensus TATA box sequence. The presence of multiple transcription start sites is consistent with the absence of a consensus TATA box sequence and the high GC content of the putative UGRP promoter (Fig. 7; Smale 1997), as well as the ability of this region to drive high level reporter gene expression (Fig. 8).

Tissue specific activity of the UGRP gene promoter

Figure 7 shows an alignment of the proximal mouse and human promoters. This alignment reveals a stretch of \(~72\%\) identity, not including spaces, over the \(-474\) to \(-1\) promoter region (Fig. 7). It is conceivable that the UGRP promoter may extend well beyond the region shown in Fig. 7;
however, a long run of T residues is located immediately adjacent to the region shown in the human, though not the mouse, UGRP gene and such a feature is not characteristic of a typical promoter. Figure 7 also shows the position of putative trans-acting factor binding sites that are
Figure 7  Sequence of the mouse and human UGRP gene promoters. The mouse and human UGRP promoter sequences are labeled relative to the predicted translation start site of UGRP, designated as +1. Putative trans-acting factor binding sites that are conserved in the mouse and human UGRP promoters, as predicted by the MatInspector sequence analysis software (Quandt et al. 1995), are boxed as are single nucleotides the sequence of which are variable (V) in the human (VH) and mouse (VM) genomes. The VH nucleotide differs between human genomic clones AC004150 and AC023855. The sequence shown is taken from AC023855 and matches that of the human UGRP gene in the Celera Discovery System database. The VM nucleotide differs between mouse genomic clone AL954730 and the mouse UGRP gene in the Celera Discovery System database. The VM nucleotide sequence of mouse genomic clone AL954730 matches that of the mouse AK002966 cDNA whereas the sequence in the mouse UGRP promoter isolated by PCR matches that of the mouse UGRP gene in the Celera Discovery System database. The latter sequence is shown. The location of primers used to amplify the mouse UGRP promoter are underlined. Met, translation start site. The experimentally determined transcription start sites in the mouse gene are indicated by black dots. The mouse UGRP cDNA in the NCBI nr database (accession number AK002966) differs from all the cDNAs in the mouse_est database, based on the available sequence, in that it contains an additional exon that is generated by splicing of the 5′ untranslated region. The location of this alternate exon, designated 1A, is shown in square brackets. The precise location of exon 1A is based on the mouse UGRP transcript sequence in the Celera database since this sequence perfectly matches that of the mouse UGRP gene. The sequence of the mouse UGRP cDNA clone present in the NCBI nr database that contains exon 1A (AK002966) has three discrepancies with the genomic sequence in this region which presumably represent sequencing/cloning errors.
conserved between the mouse and human UGRP promoters, as predicted by the MatInspector sequence analysis software (Quandt et al. 1995).

Several of these trans-acting factors including nuclear factor kappaB (NFkB) (Barnes & Karin 1997), Sp1 (Suske 1999), and E-Box binding factors (Massari & Murre 2000) have a wide tissue distribution consistent with the ubiquitous expression pattern of UGRP mRNA (Fig. 3). The computational analysis also predicts that factors which show a more restricted expression pattern such as hepatocyte nuclear factor-3 (HNF-3) (Lai et al. 1993) will bind the UGRP promoter so it is possible that they may modulate UGRP gene expression in certain cell types.

The mouse and human UGRP promoter sequences shown in Fig. 7 were ligated to the firefly luciferase reporter gene and the expression of the resulting fusion genes were analyzed by transient transfection of the muscle-derived L6 cell line, the pancreatic islet β cell-derived βTC-3 cell line, the liver-derived HepG2 cell line and the cervix-derived HeLa cell line (Fig. 8). Basal mouse and human UGRP fusion gene expression was compared with that directed by a mouse G6Pase catalytic subunit-firefly luciferase fusion gene construct containing promoter sequence from −231 to +66, a region that appears to contain all the elements required for regulation of gene expression by insulin and cAMP (Streeper et al. 1998, 2000). A similar comparison was made to a mouse IGRP-firefly luciferase fusion gene containing promoter sequence from −306 to +3, a region which incorporates all the elements required for maximal basal IGRP gene expression in hamster insulinoma tumour (HIT) cells (Ebert et al. 1999).

Figure 8 Comparison of the relative basal activities of the mouse G6Pase catalytic subunit, mouse (Mo.) IGRP and mouse and human (Hum.) UGRP gene promoters. βTC-3, HepG2, HeLa and L6 cells were transiently co-transfected, as described in Materials and methods, with an expression vector encoding Renilla luciferase and either a mouse or human UGRP-, mouse G6Pase catalytic subunit- or mouse IGRP-firefly luciferase fusion gene containing promoter sequence from −497 to −1, −474 to −1, −231 to +66 or −306 to +3 respectively. Following transfection, cells were harvested after 18–20 h in serum-free (HepG2, HeLa and L6) or serum-containing (βTC-3) medium and both firefly and Renilla luciferase activity were assayed as described in Materials and methods. Results are presented as the ratio of firefly to Renilla luciferase activity and represent the mean of three experiments±S.E.M., each using an independent preparation of all four fusion gene plasmids.
The relative basal expression of these fusion genes varied dramatically between the transfected cell types (Fig. 8). The mouse and human UGRP promoters were active in all four cell lines examined, with the human promoter more active than the mouse in each case. In contrast, the IGRP promoter was only active in βTC-3 cells whereas the G6Pase catalytic subunit promoter only directed high basal reporter gene expression in HepG2 cells. It should be noted that, in the presence of glucocorticoids and cAMP, G6Pase catalytic subunit fusion gene expression is strongly induced in hepatoma cells (Streeper et al. 1998, 2000, 2001) and exceeds that of UGRP (data not shown).

Discussion

This paper describes the identification and initial characterization of a cDNA and the gene encoding the ubiquitously expressed mouse G6Pase catalytic subunit-related protein (UGRP). The open reading frame of this UGRP cDNA encodes a protein (346 aa; Mr 38 755) that shares 36% overall identity (56% similarity) to the mouse G6Pase catalytic subunit (357 aa; Mr 40 454) and 35% overall identity (55% similarity) to mouse IGRP (355 aa; Mr 40 653) (Fig. 1). The mouse G6Pase catalytic subunit has 50% overall identity to mouse IGRP (Fig. 1). UGRP exhibits a similar predicted transmembrane topology to that of IGRP and the G6Pase catalytic subunit (data not shown); however, in contrast to the G6Pase catalytic subunit, neither UGRP nor IGRP appear to catalyze G6P hydrolysis after transfection into cell lines (Table 1). It is possible that neither IGRP nor UGRP are catalytically active but this seems improbable given the abundance of IGRP mRNA in islets (Martin et al. 2001) and the abundance of UGRP mRNA in multiple tissues (Fig. 3). In addition, although UGRP (and IGRP) lack some of the residues that are critical for hydrolysis of glucose-6-phosphate by the G6Pase catalytic subunit (Fig. 1; Chou & Mansfield 1999), all three proteins share an extended sequence motif found in bacterial vanadate-sensitive haloperoxidases and mammalian phosphatidic acid phosphatases (Hemrika et al. 1997, Stukey & Carman 1997) that constitutes the active site of these enzymes. Interestingly, EST database analyses reveal that UGRP, like the G6Pase catalytic subunit, is also expressed in rats in contrast to IGRP which is not expressed in this species (Martin et al. 2001).

The mouse UGRP and G6Pase catalytic subunit genes both map to chromosome 11 whereas the mouse IGRP gene is located on chromosome 2. Nevertheless, the overall exon/intron structures of the mouse G6Pase catalytic subunit, IGRP and UGRP genes are very similar (Tables 2 and 3), consistent with the evolutionary relationship which is evident from their coding sequences. One exception is the presence of an additional exon in the mouse (Fig. 1, Table 2) and human UGRP (Martin et al. 2002) genes that represents an apparent bifurcation of an exon by insertion of an intron, rather than the acquisition of extra coding sequence. In addition, the mouse UGRP gene (Table 2), like the human UGRP gene (Martin et al. 2002), is smaller in overall size than the G6Pase catalytic subunit and IGRP genes. A cross-species comparison of the exon/intron sizes of each of the three genes reveals a perfect conservation of exon coding sequence size between the mouse and human homologs with some variation in intron sizes (compare Table 2 and Martin et al. 2002).

RNA blot analysis showed that mouse UGRP mRNA was detected in every tissue examined with highest expression in kidney and brain (Fig. 3). Human UGRP is likewise broadly distributed (Martin et al. 2002) with highest expression in skeletal muscle, heart, thymus, brain and kidney (Martin et al. 2002). These patterns of mouse and human UGRP mRNA expression contrast markedly with those of IGRP and G6Pase catalytic subunit mRNA. The former is only found in pancreatic islets (Arden et al. 1999, Martin et al. 2001) whereas the latter is predominantly detected in liver and kidney (van de Werve et al. 2000, Van Schaftingen & Gerin 2002). The finding that the UGRP promoter was active in all four cell lines examined (Fig. 8) is consistent with its broader expression pattern based on mRNA analysis.

The sequence alignment of the mouse and human UGRP gene promoters shows that many, though not all, of the cis-acting elements identified in the human UGRP promoter are conserved in the mouse promoter (compare Fig. 7 and Martin et al. 2002). The conserved elements include E-Box motifs as well as HNF-3 and Sp1 binding sites. Although sequence alignments reveal little similarity between the UGRP and either the IGRP or G6Pase catalytic subunit promoters, it is interesting
that HNF-3 and Sp1 binding sites have been identified in all three promoters. Thus, Chou and colleagues have shown that multiple HNF-3 binding sites in the G6Pase catalytic subunit promoter are important for basal expression of that gene (Lin et al. 1997) and we have previously shown that HNF-3 is important for basal IGRP fusion gene expression (Bischof et al. 2001). Similarly, Schmoll and colleagues (Wasner et al. 2001) have shown that Sp1 is important for basal G6Pase catalytic subunit gene expression and we have previously identified an Sp1 motif in the proximal IGRP promoter region (Bischof et al. 2001). The role of HNF-3 and Sp1 in basal UGRP gene expression will be investigated in future experiments.

In contrast to the HNF-3 and Sp1 binding sites found in all three promoters, E-Box motifs are only apparent in the UGRP (Fig. 7) and IGRP promoters (Bischof et al. 2001). E-Box motifs conform to the consensus sequence CANNTG, bind to transcription factors of the basic helix-loop-helix (bHLH) family (Massari & Murre 2000) and play a role in the differentiation of multiple tissues including the nervous system, pancreas and muscle (Guillemot et al. 1993, Naya et al. 1997, Arnold & Winter 1998). We have recently shown that the high activity exhibited by the mouse IGRP promoter in islet cell lines is dependent on the integrity of two conserved E-boxes and that these motifs bind to members of the bHLH family, specifically upstream stimulatory factor (USF) and BETA2/NeuroD (Martin et al. 2003). The absence of E-Box motifs in the G6Pase catalytic subunit promoter is surprising in that glucose stimulates expression of this gene (Massillon 2001) and E-Box-like motifs mediate the action of glucose on multiple genes (Foufelle & Ferre 2002). However, it is possible that an alternate target of glucose action, such as Sp1 (Samson & Wong 2002), mediates the action of glucose on G6Pase catalytic subunit gene expression.

In summary, this paper describes the characterization of a mouse UGRP cDNA and the mouse UGRP gene and a comparison between the structures of the equivalent human molecules. Future studies will focus on defining the nature of the biochemical reaction catalyzed by the UGRP and the identification of the cis-acting elements that confer the ubiquitous expression of the UGRP gene. We anticipate that the identification of conserved cis-acting elements in the mouse and human UGRP promoters will expedite this analysis.

Acknowledgments

We thank Roland Stein, Daryl Granner, Shimon Efrat, Richard Whitesell and Jun-ichi Miyazaki for providing the HeLa and INS-1, HepG2, βTC-3, L6, and Min 6 cell lines respectively. We also thank Dr David Brindley for testing phosphatidic acid as a substrate for UGRP. Research in the laboratory of R O'B was supported by NIH grant DK56374 and the Vanderbilt Diabetes Core laboratory (P60 DK20593). Research in the laboratory of J C H was supported by the American Diabetes Association (9901–116) and the Barbara Davis Center Diabetes and Endocrinology Research Center (P30 DK57516). J N B was supported by the Vanderbilt Molecular Endocrinology Training Program (5 T32 DK07563). C C M was supported by the Vanderbilt Viruses, Nucleic Acids and Cancer Training Program (5T32 CA09385). Where indicated in the text, sequence data was generated and/or analyzed through use of the Celera Discovery System and Celera’s associated databases.

References


www.endocrinology.org


Consoli A 1992 Role of liver in pathophysiology of NIDDM. *Diabetes Metabolism Research and Reviews* **8** 832–841.


factor-binding protein 1 genes. Molecular and Cellular Biology 15 1747–1758.


Shelly LL, Lei KJ, Pan CJ, Sakata SF, Ruppert S, Schutz G & Chou JY 1993 Isolation of the gene for murine glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1A. Journal of Biological Chemistry 268 21482–21485.


Received 24 April 2003
Accepted 31 October 2003
Made available online as an Accepted Preprint 12 November 2003