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

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, Vanderbilt University Medical School, Nashville, Tennessee 37232, USA
¹Barbara Davis Center for Childhood Diabetes, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

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 human cDNA and gene encoding a ubiquitously expressed G6Pase catalytic subunit-related protein (UGRP). The ORF of this UGRP cDNA encodes a protein (346 amino acids (aa); $M_r$ 38 709) which shares 36% overall identity to the human G6Pase catalytic subunit (357 aa; $M_r$ 40 487). 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. UGRP mRNA was detected by RNA blot analysis in every tissue examined with the highest expression in muscle. Database analysis showed that the human UGRP gene is composed of six exons, spans ~5·4 kbp of genomic DNA and is located on chromosome 17q21 with the G6Pase catalytic subunit gene. The UGRP gene transcription start sites were mapped by primer extension analysis, and the activity of the 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 (2002) 29, 205–222

Introduction

Glucose-6-phosphatase (G6Pase) is located in the endoplasmic reticulum (ER) and catalyzes the terminal step of the gluconeogenic and glycogenolytic pathways in liver, the hydrolysis of glucose-6-phosphate (G6P) to glucose and phosphate. The enzyme is thought to exist as a multi-component system; however, the exact number of components and their stoichiometry are unclear (Foster et al. 1997, Mithieux 1997, van de Werve et al. 2000, Van Schaftingen & Gerin 2002). Two components of the system have been identified to date, namely the catalytic subunit (Lei et al. 1993) and a G6P transporter (Gerin et al. 1997). Both are integral membrane proteins that span the ER membrane multiple times. Studies using microsomes suggest that the G6Pase catalytic subunit has its catalytic site oriented towards the lumen of the ER (Foster et al. 1997, Mithieux 1997, van de Werve et al. 2000, Van Schaftingen & Gerin 2002). In this model, the G6P transporter serves to deliver cytosolically generated G6P to the active site of the catalytic subunit. This model also requires an inorganic phosphate and glucose transporter to shuttle the reaction products back to the cytosol; however, neither of these components has been identified (Foster et al. 1997, Mithieux 1997). An alternative model, based on rapid kinetic data, places the catalytic site within the membrane and
ascribes both a transport function and catalytic activity to the catalytic subunit (Foster et al. 1997, Mithieux 1997). However, the validity of this rapid kinetic data has been challenged (Arion et al. 1998) and recent data favor the substrate-transport model of G6Pase (Gerin et al. 2001).

Inactivating mutations within the G6Pase catalytic subunit and G6P transporter cause glycogen storage disease (GSD) types 1a and 1b respectively (Chou & Mansfield 1999), diseases characterized by fasting hypoglycemia, hepatic glycogen accumulation and hepatomegaly. Increased hepatic G6Pase activity, on the other hand, contributes to the augmented hepatic glucose production and the hyperglycemia that characterizes both type 1 and 2 diabetes (Consoli 1992, Granner & O’Brien 1992, Cline et al. 1994, O’Brien & Granner 1996). Accordingly, overexpression of the G6Pase catalytic subunit or the G6P transporter in hepatocytes (An et al. 2001, Seoane et al. 1997) or in vivo (Trinh et al. 1998) using recombinant adenovirus virus enhances the rate of G6P hydrolysis.

While G6Pase activity is highest in liver, it is also detected in kidney, intestine and islets (Foster et al. 1997, Mithieux 1997). The actual contribution of the kidney, relative to liver, to the overall rate of gluconeogenesis is the subject of continuing debate (Stumvoll et al. 1997, Cherrington 1999) and recent data suggest that intestine may also play a major role (Mithieux 2001). Although G6Pase activity is detectable in islets it displays distinct kinetic behavior and inhibitor profiles compared to 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). Studies on G6P transport in fibroblast microsomes isolated from patients with GSD type 1b likewise demonstrate the existence of a second G6P transporter (Leuzzi et al. 2001).

We have recently identified a novel cDNA (Arden et al. 1999) and gene (Ebert et al. 1999) which encodes an islet-specific G6Pase catalytic subunit-related protein (IGRP) which is similar in size (355 amino acids (aa); Mr 40 552), topology and sequence (~50% identity at the amino acid level) to the G6Pase catalytic subunit. Overexpression of IGRP in tissue culture cells, however, 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. Since the function of IGRP is uncertain we have continued to search for an additional G6Pase catalytic subunit isoform that might account for the unique kinetics of the G6Pase reaction in islets. In the course of this search we have identified a cDNA and gene encoding a ubiquitously expressed G6Pase catalytic subunit-related protein (UGRP), a third member of the G6Pase gene family. UGRP is similar in size, topology and sequence to the G6Pase catalytic subunit but, as with IGRP, overexpression of UGRP in tissue culture cell studies does not increase G6P hydrolysis. However, the exon/intron structures of all three genes are similar, suggesting that they are evolutionarily related.

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 analytical grade and purchased from Sigma Chemical Company (St Louis, MO, USA). Human muscle RNA was purchased from Ambion, Inc. (Austin, TX, 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 spin columns (Bio-Rad, Hercules, CA, USA). DNA sequencing was performed using the USB Sequenase kit or by automated sequencing using an ABI 310 DNA analyzer (Applied Biosystems). Because transcription of the UGRP gene initiates at multiple start sites (see Fig. 4), 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).
Identification of human UGRP cDNA and genomic clones

A BLAST search of the NCBI nr database (http://www.ncbi.nlm.nih.gov) using the human IGRP peptide sequence as the query identified a novel clone, Accession number AAH02494, related to the IGRP and G6Pase catalytic subunit. A full-length cDNA encoding this novel peptide, ligated into the pOTB7 vector as an EcoRI-XhoI fragment (Accession number BC002494; I.M.A.G.E. Clone ID 3050476), was purchased from ATCC for use in RNA blotting and the construction of a UGRP expression vector (see below). A BLAST search of the NCBI nr and htgs databases using this human UGRP cDNA sequence as the query identified a PAC clone (Accession number AC004150) and a BAC clone (Accession number AC023855) respectively, both of which contained the entire human UGRP gene as a contiguous sequence.

Endogenous UGRP expression analysis

The tissue-specific expression of the UGRP gene was analyzed by RNA blotting using commercially available blots (MTN #7760-1 and MTN #7780-1; Clontech, Palo Alto, CA, USA) prepared with 1–2 µg samples of poly A+ mRNA from various human tissues. Blots were hybridized for 1 h at 65 °C in Clontech ExpressHybTM Hybridization Solution with 100 µg salmon testis DNA and a 32P-radiolabeled randomly primed probe corresponding to the full-length UGRP ORF (ORF). This probe was isolated from the pOTB7 vector described above as an EcoRI-XhoI fragment and then 25 ng was labeled by random oligonucleotide priming with [α-32P]dATP using the Stratagene Prime-It II random primer labeling kit (Stratagene, San Diego, CA, USA) according to the manufacturer’s instructions (final specific radioactivity ~0.5 Ci/µmol). After hybridization, blots were washed in 2 × SSC, 0.05% SDS at room temperature for 30 min, then in 0.1 × SSC, 0.1% SDS at 42 °C for up to 2 h prior to visualization by autoradiography. Blots were then stripped by heating for 10 min in 0.5% SDS at 95 °C and then re-probed with a labeled human β actin cDNA (Clontech) to confirm similar mRNA loading levels in each lane.

Primer extension analysis

Human muscle total RNA was purchased from Ambion Inc. A 24 bp primer (5’-GATGCCGCGCCCAAGCGTGGAATC-3’), complementary to 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 to 50 µg total RNA for 1 h at 60 °C and then primer extension was performed as previously described (Ebert et al. 1999).

Cell culture

Islet-derived βTC-3 cells, liver-derived HepG2 cells, cervix-derived HeLa cells and kidney-derived COS 7 cells were passaged as subconfluent cultures in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 100 U/ml penicillin and 100 µg/ml streptomycin. Cell cultures were also supplemented with either 2.5% (v/v) fetal bovine serum and 15% (v/v) horse serum (βTC-3 cells), 2.5% (v/v) newborn calf serum, 2.5% (v/v) fetal bovine serum and 5% (v/v) Nu serum IV (Becton Dickinson, Bedford, MA, USA) (HepG2 cells), 10% calf serum (HeLa cells) or 10% (v/v) fetal bovine serum (COS 7 cells).

Fusion gene plasmid construction and analysis

The construction of mouse G6Pase catalytic subunit-chloramphenicol acetyltransferase (CAT) and IGRP–CAT fusion genes containing promoter sequence, relative to the transcription start site, from –231 to +66 and –306 to +3 respectively, in the pCAT(An) expression vector (Jacoby et al. 1989), have previously been described (Ebert et al. 1999, Streepen et al. 1998). For the experiments described in this paper CAT was replaced with the more sensitive firefly luciferase reporter. This was achieved by re-isolating the G6Pase catalytic subunit and IGRP promoters, as XhoI-XhoI and BamHI-XhoI fragments respectively, from the plasmids described above and ligation into the pGL3-Mod vector. pGL3-Mod was generated by replacing the polynucleotide in the pGL3-Basic vector (Promega, Madison, WI, USA) with a polynucleotide sequence of the UGRP gene.

www.endocrinology.org
containing the following restriction endonuclease recognition sites: KpnI, BamHI, HindIII, XbaI, XhoI and BglII.

A clone containing the entire UGRP gene (see above; Accession number AC023855) was purchased from BACPAC Resources, Children's Hospital Oakland Research Institute, Oakland, CA, USA). This clone was digested with EcoRI to isolate a 1289 bp fragment, containing the UGRP promoter and exon 1, which was then ligated into an EcoRI-digested pGEM7 vector (Promega). A fragment of the human UGRP promoter was then generated in a PCR reaction using this plasmid as the template in conjunction with the following primers: 5'-CCCAGCTTGATTGTGCTGGAGAAGAGAG-3' and 5'-GAAGATCTGGCCGGCCAGCTGACCAGGCC-3'; HindIII and BglII cloning sites are underlined. These primers represent the putative UGRP promoter sequences from −474 to −450 and −24 to +1, relative to the transcription start site, respectively (see Fig. 5). The PCR fragment generated was digested with HindIII and BglII and 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 hg16 database for Accession number AC023855. The HindIII and BglII UGRP promoter fragment was then re-isolated from the pSP72 plasmid and ligated 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 either 0.15 µg or 1.2 ng of a SV40-renilla luciferase plasmid (Promega) as previously described (Dickens et al. 1998). To correct for variations in transfection efficiency, the results are expressed as a ratio of firefly to renilla luciferase activity. Fusion gene expression was assessed using three independent preparations of each plasmid, transfected in triplicate. Background firefly luciferase activity obtained in transfections with the pGL3-Mod vector, which represents read-through transcription, was subtracted in each case.

**UGRP protein expression by in vitro translation and cellular transfection**

The human UGRP ORF was amplified from the above cDNA construct in pOTB7 using the forward primer, 5'-CACCATTGGAGTTCACCCGCGTG-3' and, as a reverse primer, either 5'-TCAGGAGATGAGGCTGCGGCG-3' or 5'-AGATGGAGGCGGGGT-3'. The forward primer incorporates a modified Kozak sequence (Kozak 1991) 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 of both constructs showed that the inserted ORF was identical to the initial I.M.A.G.E. Clone ID 3050476. 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 a 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 human UGRP sequences. The 35S-methionine-labeled products were separated by SDS-PAGE on 12.5% (w/v) acrylamide 0.125% N, N' bis-acrylamide gels and visualized by direct phosphorimaging (STORM phosphorimager; Molecular Dynamics, Sunnyvale, CA, USA).

The mouse G6Pase catalytic subunit, mouse IGRP and human UGRP proteins were expressed by transient transfection of COS 7 cells using a commercial lipofection reagent (Fugene; Boehringer Mannheim). 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 5 x 10^5 COS 7 cells which had been
freshly plated in 4 ml DMEM containing 10% (v/v) fetal calf serum and antibiotics in 6 cm Petri dishes. Cells were harvested after 48 h using a non-enzymatic procedure (GIBCO cell dissociation buffer; Life Technologies, Rockville, MD, USA), 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 MES (2-(N-Morpholino)-ethanesulfonic acid) K+, 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 dimethyl glutarate buffer (pH 6·5) containing 10 mM NaK tartrate, 10 mM EDTA and 10 mM G6P as previously described (Arden et al. 1999). β-Galactosidase activity in the supernatant from the homogenate was assayed using a previously described spectrophotometric assay (Arden et al. 1999).

To demonstrate that the mouse G6Pase catalytic subunit, mouse IGRP and human UGRP are expressed at similar levels in COS cells, 100 µg cell extract containing the pcDNA3·1D/V5 His variants of these proteins was electrophoresed on 12·5% SDS-polyacrylamide gels and the proteins transferred to nitrocellulose as previously described (Hutton et al. 1988). Immunoblot analysis was then performed, again as previously described (Hutton et al. 1988), using a horseradish peroxidase-conjugated anti-V5 antibody (Invitrogen).

Results

Identification and sequence analysis of UGRP

A BLAST search of the NCBI nr database using the human IGRP peptide sequence as the query identified a novel homologous peptide sequence, Accession number AAH02494, related to the G6Pase catalytic subunit. The sequence of this novel peptide, designated UGRP, was derived from the sequence of an I.M.A.G.E. clone ID 3050476 (Accession number BC002494). This 1420 bp cDNA clone contains a poly A+ tract at the 3’ end and appears to contain the entire UGRP ORF. Thus, analysis of UGRP genomic sequence (see below) identifies an in-frame stop codon located 606 bp 5’ of the putative initiation methionine. This 606 bp fragment contains one other ATG codon but this is out-of-frame with the predicted UGRP peptide sequence. 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 UGRP cDNA does not exist. Indeed, this conclusion is supported by an analysis of the NCBI human_est database that identifies multiple UGRP cDNAs with variable 5’ end points. This analysis of other UGRP cDNA clones also suggests that the first 7 bp of I.M.A.G.E. clone 3050476 represents non-UGRP sequence and therefore presumably represents a cloning artifact. If correct, this clone contains 36 bp of 5’ untranslated leader sequence.

The deduced human UGRP ORF (Fig. 1) encodes a 346 amino acid protein (Mw 38 709, pI 8·57) of a generally hydrophobic character (239/346 residues). The hydrophobic amino acids are arranged in nine major stretches, eight of which are predicted to be able to span a phospholipid bilayer as an alpha helix (TMAP; Milpetz et al. (1995); see //condor.bcm.tmc.edu; data not shown). Six of these stretches, however, contain charged amino acid(s) (aa): the sequence from aa 25–49, 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 115–135, 172–188 and 196–222. The UGRP ORF sequence can be aligned with residues 7–357 of the human G6Pase catalytic subunit (357 aa) and residues 1–355 of human IGRP (355 aa), the gaps occurring in non-conserved regions between the latter molecules. The sequence is 36% identical (64% similarity) to the human G6Pase catalytic subunit and is 37% identical (76% similarity) to human IGRP with the homology extending over the full length of the molecules including the putative transmembrane domains. IGRP and the G6Pase catalytic subunit are more closely related (50% identity; 75% similarity) (Fig. 1). The conservation of these hydrophobic segments and the charged residues within them suggest that they
Figure 1 Alignment of the deduced peptide sequences of the human G6Pase catalytic subunit, IGRP and UGRP.

The alignment of the peptide sequence of human UGRP (Accession number BC002494) with that of the human G6Pase catalytic subunit (Accession number U01120) and human IGRP (Accession number NM021176) is shown. Identities are indicated by solid circles and similarities by vertical bars. Residues defined as being of key catalytic importance in haloperoxidases and related phosphatases are shaded blue. Inactivating point mutation that abolish (black letters) or reduce (red letters) G6Pase catalytic subunit activity and give rise to GSD type 1a (Chou & Mansfield 1999) are shown in single letter code above the UGRP sequence.
may have a function other than simple membrane spanning. Unlike the G6Pase catalytic subunit and IGRP, UGRP contains no consensus sites for NH₂-linked glycosylation (Jackson et al. 1990) and the protein does not share the COOH-terminal consensus sequence (KKXX) that characterizes many ER-resident transmembrane proteins (Kornfeld & Kornfeld 1985).

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 human UGRP are Lys72, Arg Pro50- Pro Ser Gly His814- 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 six positions that differ between G6Pase catalytic subunit and IGRP; viz T16→I12, Q54→N50, A124→S120, L211→K209, S298→C296 and L343→V343. Among G6Pase catalytic subunit sequence from other species, position T16 is A in fish, Q54 is R in dog and fish, A124 is always conserved and often in the sequence GAAG rather than GTAG (c.f. GASC in IGRP), L211 is uniformly conserved as is L345. There are 16 positions that differ between the G6Pase catalytic subunit and UGRP that are inactivating mutations in G6Pase catalytic subunit; viz T16→A12, Q24→S50, G60→T144, T108→S103, G122→I117, A124→G119, W156→L147, G222→L212, T253→I245, N264→D254, I265→S253, R295→K283, S298→C286, V338→I325, I341→N328 and L345→V332.

Translation and enzymatic activity of human UGRP

In vitro translation of a human UGRP cRNA generated a protein of 34 kDa (39 kDa with V5 His COOH terminal extension), a size somewhat smaller than the predicted Mᵣ (38 709 Da) (Fig. 2). The size discrepancy between the predicted and observed Mᵣ of human UGRP probably relates to the hydrophobic nature of the protein. The slower electrophoretic mobility of UGRP (34 kDa) compared with mouse IGRP (40 kDa) and the mouse G6Pase catalytic subunit (37 kDa) is conceivably related to a combination of the length of the ORF (346, 355 and 357 aa respectively) and more acidic nature (predicted pI 8.57, 8.72 and 9.22, respectively).

Enzyme activity studies were performed by measurements of G6P hydrolytic rates in COS 7 cells transiently transfected with cytomegalovirus (CMV) promoter-driven constructs containing the mouse G6Pase catalytic subunit, mouse IGRP or human UGRP ORFs linked to a human growth hormone 3’ untranslated region and polyadenylation signal (pcDNA 3·1D TOPO vector). Constructs were generated which either terminated at the native stop codon or encoded a protein, designated V5 His (Fig. 2 and 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 enzymatic 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 the V5 His variants of the mouse G6Pase catalytic subunit, mouse IGRP and human UGRP were all expressed at similar levels in COS cells (Fig. 2B).

Transfection with the G6Pase catalytic subunit construct resulted in an ~80-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 human UGRP or IGRP did not produce 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; 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 either the G6Pase catalytic subunit, IGRP or UGRP encoding plasmids, thus indicating that neither cell viability nor biosynthetic capacity was adversely affected by introduction of these constructs (Table 1). The rate of hydrolysis of a series of low Mᵣ
phosphorylated substrates (mannose-6-phosphate, pyrophosphate, phosphoenolpyruvate, ribose-5-phosphate, ribulose-5-phosphate, 6-phosphogluconate, fructose-1-phosphate, fructose-6-phosphate, fructose-1,6-bisphosphate and glucosamine-6-phosphate) that can be hydrolyzed by the G6Pase catalytic subunit (Arden et al. 1999) was not significantly increased by the introduction of UGRP (data not shown). The rate of hydrolysis of the generic phosphatase substrate, p-nitrophenol phosphate (pNPP), was also unaltered in human UGRP-transfected COS 7 cells (data not shown), yet it proved a good substrate for the G6Pase catalytic subunit in transfected cells (Arden et al. 1999). Enzyme assays performed at pH 5 or pH 8 in the presence or absence of divalent cations (0.2–2 mM Mg$^{2+}$ or Ca$^{2+}$) or thiol reagents (5 mM dithiothreitol) did not reveal any hydrolytic activity associated with UGRP. The addition of a COOH terminal V5 epitope/His tag that could potentially affect the localization of the G6Pase catalytic subunit did not impede its hydrolytic activity significantly or affect G6P hydrolysis in IGRP- and UGRP-transfected cells (Table 1).

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

RNA blot analyses of human tissue poly A$^+$ mRNA using the human cDNA described above as the labeled probe showed the presence of a single ~1500 nucleotide hybridizing species in 13 major

---

**Figure 2** *In vitro* and *in situ* expression of the mouse G6Pase catalytic subunit, mouse IGRP and human UGRP. (A) T7 polymerase-derived transcripts from cloned mouse G6Pase catalytic subunit, mouse IGRP and human UGRP pcDNA3.1D constructs with or without the V5 His$_6$ COOH terminal extension (V5 His) were translated *in vitro* in the presence of $^{35}$S methionine using rabbit reticulocyte lysate and analyzed by SDS-PAGE as described in Materials and methods. (B) Plasmids encoding the V5 His variants of the mouse G6Pase catalytic subunit, mouse IGRP and human UGRP were transiently transfected into COS cells and protein expression was then estimated by immunoblot analysis as described in Materials and methods.
human tissues. The highest level of UGRP expression was observed in skeletal muscle and intermediate levels were detected in heart, brain, placenta, kidney, colon, thymus, spleen and pancreas (Fig. 3). The lowest levels of UGRP expression were found in liver, lung, small intestine and peripheral blood leukocytes (PBL); expression in the latter two tissues was apparent after longer exposure of the film (data not shown).

The calculated size of human UGRP mRNA without a poly A+ tail is 1339 bp (Table 2), a size that is consistent with the 1500 nucleotide hybridizing species seen on RNA blots (Fig. 3) given that the UGRP cDNAs reported in the NCBI human_est database contain variable 5′ untranslated leader sequences up to 250 bp in length. Since the RNA loaded on these blots was from different individuals it is not possible to make a precise quantitative assessment based on these data, especially as the relative UGRP expression levels varied between the two MTN blots examined (Fig. 3). Nevertheless, the broad tissue distribution contrasts with that of the G6Pase catalytic subunit and IGRP whose distribution are highly restricted (Foster et al. 1997, Mithieux 1997, Martin et al. 2001).
arisen from insertion of an intron in the UGRP gene within the equivalent of exon 5 of the G6Pase catalytic subunit gene (or conversely its excision from an ancestral G6Pase catalytic subunit gene). The other notable difference is that the 3′ untranslated region in the human UGRP gene that is contained within exon 6 (301 bp with a consensus polyadenylation signal sequence (Colgan & Manley 1997), 14 bp 5′ of the known poly A+ addition site) is shorter than that of either the G6Pase catalytic

![Figure 3](https://example.com/image.png)

**Figure 3** Analysis of human UGRP mRNA expression by RNA blotting. RNA blotting analysis was performed on the indicated ∼1–2 µg samples of poly A+ mRNA. Clontech MTN blot #7780-1 (A) and MTN #7760-1 (B) blots were hybridized with a labeled human UGRP cDNA spanning the ORF prior to washing and visualization by autoradiography as described in Materials and methods. Blots were then stripped and re-probed with a labeled human β-actin cDNA to confirm similar mRNA loading levels in each lane. Representative experiments are shown.
subunit or IGRP genes (both >1500 bp contained within exon 5). The exact size of exon 5 in the human G6Pase catalytic subunit and IGRP genes are unknown since poly A+ cDNAs have not been isolated (Table 2). The sizes of the four introns are similar between the human G6Pase catalytic subunit and IGRP with the exception of intron A which is smaller in the IGRP gene. In contrast, the sizes of introns B–D are considerably smaller in the UGRP gene than the G6Pase catalytic subunit gene. Consequently, the G6Pase catalytic subunit gene spans ~12·1 kbp compared with ~8·7 kbp for IGRP and ~5·4 kbp for UGRP (Table 2). The exon/intron boundaries are generally well conserved between the human G6Pase catalytic subunit, UGRP and IGRP genes (Table 3 and Martin et al. 2001) 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 sometimes 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 human UGRP gene

The PAC clone (Accession number AC004150) and BAC clone (Accession number AC023855), containing the entire human UGRP gene, that were identified by a BLAST search of the NCBI nr and htgs databases respectively, using the human UGRP cDNA sequence as the query, are both reported to represent human chromosome 17 sequence. The genes encoding the human G6Pase catalytic subunit and IGRP have been mapped to chromosome 17 (Lei et al. 1994) and chromosome 2 (Martin et al. 2001) respectively. Further analysis using the Celera Discovery System and Celera’s associated databases revealed that the genes encoding the human G6Pase catalytic subunit and UGRP both map to chromosome 17q21. In contrast, the human G6P transporter gene has been mapped to chromosome 11q23 (Veiga-da-Cunha et al. 1998).

Determination of the UGRP gene transcription start site

To determine the transcription start site of the UGRP gene, a primer extension analysis was performed using human muscle RNA (Fig. 4). The primer for this analysis was designed to be complementary to the UGRP coding sequence between +3 and +26. Multiple transcription start sites were detected using this primer, the most prominent of which were clustered over two separate regions of 11 and 4 bp located

Table 2 Structure of the human G6Pase catalytic subunit, IGRP and UGRP genes. The human G6Pase catalytic subunit and human IGRP exon sizes are from Lei et al. (1993) and Martin et al. (2001) respectively. The human UGRP gene intron sizes are from Martin et al. (2001). The human UGRP gene exon and intron sizes as well as the G6Pase catalytic subunit intron sizes were determined as described in the Results. Poly A+ -containing cDNAs encoding the human G6Pase catalytic subunit and human IGRP have not been described so the precise size of exon 5 in these genes remains to be determined. The length of the coding sequence in exon 1 (218 bp) is identical between all three genes.

<table>
<thead>
<tr>
<th>Exon number</th>
<th>G6Pase</th>
<th>IGRP</th>
<th>UGRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>309</td>
<td>260</td>
<td>218</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>110</td>
<td>107</td>
</tr>
<tr>
<td>3</td>
<td>106</td>
<td>112</td>
<td>91</td>
</tr>
<tr>
<td>4</td>
<td>116</td>
<td>116</td>
<td>119</td>
</tr>
<tr>
<td>5</td>
<td>&gt;2000</td>
<td>~2400</td>
<td>142</td>
</tr>
<tr>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
<td>662</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intron number</th>
<th>G6Pase</th>
<th>IGRP</th>
<th>UGRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2829</td>
<td>895</td>
<td>2974</td>
</tr>
<tr>
<td>B</td>
<td>3490</td>
<td>1950</td>
<td>413</td>
</tr>
<tr>
<td>C</td>
<td>1673</td>
<td>2046</td>
<td>198</td>
</tr>
<tr>
<td>D</td>
<td>1496</td>
<td>789</td>
<td>222</td>
</tr>
<tr>
<td>E</td>
<td>N/A</td>
<td>N/A</td>
<td>228</td>
</tr>
</tbody>
</table>
between −21 and −24 and −34 and −44 respectively (Fig. 4). Because transcription of the UGRP gene initiates at multiple start sites, all UGRP DNA sequences are numbered relative to the predicted translation start site (Fig. 5). This result 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. 5; see Smale 1997), as well as the ability of this sequence to drive high level reporter gene expression (Fig. 6).

Tissue-specific activity of the UGRP gene promoter

Figure 5 shows the position of putative trans-acting factor binding sites in the UGRP promoter, as predicted by the MatInspector sequence analysis software (Quandt et al. 1995). Several of these trans-acting factors, including Egr-1 (Gashler & Sukhatme 1995), NFκB (Barnes & Karin 1997), Sp1 (Suske 1999), p53 (Vogelstein et al. 2000) 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 more restricted expression patterns such as HNF-3 (Lai et al. 1993) and Brn-2 (Hara et al. 1992) will bind the UGRP promoter so it is possible that they may modulate UGRP gene expression in certain cell types. It is also conceivable that the UGRP promoter may extend well beyond the −474 to −1 region shown in Fig. 5; however, a long run of T residues is located immediately adjacent to the region shown and such a feature is not characteristic of a typical promoter.

The putative UGRP promoter sequence shown in Fig. 5 was ligated to the firefly luciferase reporter gene and expression of the resulting fusion gene was analyzed by transient transfection of the islet-derived βTC-3 cell line, the liver-derived HepG2 cell line and the cervix-derived HeLa cell line (Fig. 6). Basal UGRP fusion gene expression was compared with that of a G6Pase catalytic subunit–firefly luciferase fusion gene construct.
containing promoter sequence from −231 to +66, a region which contains all the elements required for regulation of gene expression by insulin and cAMP (Streeper et al. 1998, 2000, 2001). A similar comparison was made with an 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 tumor cells (Ebert et al. 1999). The relative basal expression of these fusion genes varied dramatically between the transfected cell types (Fig. 6). The UGRP promoter was active in all three cell lines examined. In contrast, the IGRP promoter was only active in βTC-3 cells whereas the G6Pase catalytic subunit promoter directed high reporter gene expression in HepG2 cells.

Discussion

This paper describes the identification and initial characterization of a cDNA and the gene encoding UGRP. The ORF of this UGRP cDNA encodes a protein (346 aa; 38 709) that has 36% overall identity to the human G6Pase catalytic subunit (357 aa; 40 487) and 37% overall identity to human IGRP (355 aa; 40 552) (Fig. 1). The human G6Pase catalytic subunit has 50% overall identity to human IGRP (Fig. 1). UGRP and IGRP both exhibit a similar predicted transmembrane topology with 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). At present we cannot rule out the possibility that neither IGRP nor UGRP are catalytically active. The latter seems improbable given the abundance of IGRP mRNA in islets (Martin et al. 2001), the abundance of UGRP mRNA in muscle (Fig. 3) and the conservation of consensus phosphatase sequences (Fig. 1; see Chou & Mansfield 1999). UGRP, unlike the G6Pase catalytic subunit and IGRP, lacks the COOH-terminal consensus sequence (KKXX) characteristic of an ER-resident transmembrane protein (Jackson et al. 1990). However, this motif is not essential for expression of G6Pase catalytic subunit enzymatic activity (J C Hutton, unpublished observations) and in any case does not preclude localization of UGRP to the ER compartment (Jackson et al. 1990).

The exon/intron structures of the G6Pase catalytic subunit, IGRP and UGRP genes are very similar (Tables 2 and 3), consistent with the evolutionarily relationship which is evident from their coding sequences. The UGRP and G6Pase catalytic subunit genes both map to chromosome 17q21 whereas the IGRP gene is located on chromosome 2. The UGRP gene is smaller in overall size and contains an additional exon as the result of an apparent bifurcation of an exon by insertion of an intron, rather than the acquisition of
an additional coding sequence (Fig. 1 and Table 2). Major differences, however, are apparent at the level of its promoter structure; whereas the mouse IGRP and mouse G6Pase catalytic subunit promoter sequences share $\sim$50% homology, not including spaces, over the IGRP $-252$ to $+1$ region (Ebert et al. 1999), the sequence of the UGRP promoter does not align with that of either the G6Pase catalytic subunit or IGRP promoters as determined using the IntelliGenetics, Inc. IFIND program. Interestingly, subsequent fusion gene analyses have suggested that, despite this clear homology between the IGRP and G6Pase catalytic subunit promoters, it is not of functional significance (Ebert et al. 1999, Martin et al. 2001). Thus, the IGRP promoter is completely inactive in HepG2 cells (Fig. 6; Ebert et al. 1999) and the G6Pase catalytic subunit promoter is inactive in $\beta$TC-3 cells (Fig. 6).

RNA blot analysis showed that UGRP mRNA was detected in every tissue examined with highest expression in muscle (Fig. 3). In contrast, IGRP mRNA is only detected in pancreatic islets (Arden et al. 1999, Martin et al. 2001) whereas the G6Pase

![Figure 5](image-url) Sequence of the human UGRP gene promoter. The human UGRP promoter sequence is labeled relative to the predicted translation start site of UGRP, designated as $+1$. Putative trans-acting factor binding sites in the UGRP promoter, as predicted by the MatInspector sequence analysis software (Quandt et al. 1995), are boxed as is a single nucleotide, labeled as variable (V), the sequence of which differs between 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 location of primers used to amplify the UGRP promoter are underlined. Met, translation start site. The experimentally determined transcription start sites are indicated by asterisks.
Figure 6 Comparison of the relative basal activities of the human G6Pase catalytic subunit, IGRP and UGRP gene promoters. βTC-3, HepG2 and HeLa cells were transiently co-transfected, as described in Materials and methods, with an expression vector encoding renilla luciferase and either a G6Pase catalytic subunit–, IGRP– or UGRP–firefly luciferase fusion gene containing promoter sequence from −231 to +66, −306 to +3 or −474 to −1 respectively. Following transfection, cells were harvested after 18–20 h 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 three fusion gene plasmids.
catalytic subunit gene is predominantly expressed in liver and kidney (Foster et al. 1997, Mithieux 1997). These different expression patterns for the endogenous G6Pase catalytic subunit, IGRP and UGRP genes are reflected in the tissue-specific activity of the three gene promoters as assessed using fusion gene constructs (Fig. 6). Thus, in marked contrast to the IGRP and G6Pase catalytic subunit gene promoters, the UGRP promoter was active in all three cell lines examined (Fig. 6).

Although not apparent through sequence alignments, all three promoters contain putative binding sites for the forkhead transcription factor HNF-3, now known as Foxa2 (Fig. 5; Lin et al. 1997, Bischof et al. 2001). We have previously shown that HNF-3 is important for basal IGRP fusion gene expression (Bischof et al. 2001) and Lin et al. (1997) have shown that multiple HNF-3 binding sites in the G6Pase catalytic subunit promoter are important for basal expression of that gene. In the G6Pase catalytic subunit promoter, HNF-3 may also act as an accessory factor to enhance the stimulatory effect of cAMP (Streeper et al. 2000). The role of HNF-3, as well as the other factors shown in Fig. 5, in basal UGRP gene expression will be investigated in future experiments.

The UGRP promoter contains three E-Box motifs that conform to the consensus sequence CANNTG (Fig. 6; Massari & Murre 2000). These promoter elements bind to transcription factors of the basic helix-loop-helix (bHLH) family 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). The connection between bHLH proteins and development derives primarily from studies in muscle which have shown that four bHLH proteins (MyoD, myogenin, Myf5, and MRF4) are sufficient to stimulate differentiation when overexpressed individually in cell culture (Olson 1990). Examination of expression profiles and gene deletion experiments in mice have extended these findings by showing that these proteins are important for specific stages of muscle differentiation, although they seem to exhibit some functional redundancy (Arnold & Winter 1998). bHLH proteins bind to E-Box elements as heterodimers of tissue-specific proteins and ubiquitously expressed proteins such as E12/47 or HEB (Murre et al. 1989, Lassar et al. 1991). Specifically, in muscle, it is believed that a heterodimer of MyoD and E47 can stimulate transcription of genes by synergizing with other factors, most notably MEF2C (Molkentin et al. 1995). This co-operativity leads to the recruitment of the co-activator p300/CBP which makes both direct contacts with the basal transcriptional machinery and is associated with histone acetyltransferase activity, an activity thought to relieve the transcriptional repression imposed by chromatin (Sartorelli et al. 1997, Struhl 1998). Given the known importance of E-Box motifs and the high amounts of UGRP mRNA in muscle it will be interesting to determine the functional importance of these elements in the UGRP promoter and to establish if any of the known myogenic bHLH proteins bind to them.

In a similar fashion we have recently found that the high activity exhibited by the 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 USF and BETA2/NeuroD (C C Martin & R M O’Brien, unpublished observations).

In summary, this paper describes the characterization of a human UGRP cDNA, the human UGRP gene and the initial characterization of the UGRP gene promoter. Future studies will focus on the nature of the biochemical reaction catalyzed by the UGRP, the identification of the cis-acting elements that confer the ubiquitous expression of the UGRP gene, as well as addressing the question as to whether UGRP expression is regulated in response to hormones or nutrients.

Acknowledgements

We thank Roland Stein, Daryl Granner and Shimon Efrat for providing the HeLa, HepG2 and βTC-3 cell lines respectively. Research in the laboratory of R M 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). C C M is supported by the Vanderbilt Viruses, Nucleic Acids and Cancer Training Program (5T32 CA09385-18). Where indicated in the text, sequence data were generated and/or analyzed through the use of the Celera Discovery System and Celera’s associated databases.
References


Consoli A 1992 Role of liver in pathophysiology of NIDDM. Diabetes Care 15 1430–1441.


Sartorelli V, Huang J, Hamamori Y & Kedes L 1997 Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Molecular and Cellular Biology* **17** 1010–1026.


Received in final form 12 June 2002
Accepted 19 June 2002