

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

(Requests for offprints should be addressed to R M O'Brien, Department of Molecular Physiology and Biophysics, 761 MRB II, Vanderbilt University Medical School, Nashville, Tennessee 37232-0615, USA; Email: richard.obrien@mcm.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 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 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); M_r 40 552), topology and sequence (\sim 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 - 32 P]dATP (>3000 Ci/mmol) and [γ - 32 P]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 a *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 a *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/nmol). 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'-GATGCCC GCGCCAGCGTGGACTC-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 β TC-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 (β TC-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, Streeper *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 polylinker in the pGL3-Basic vector (Promega, Madison, WI, USA) with a polylinker

containing the following restriction endonuclease recognition sites: *Kpn*I, *Bam*HI, *Hind*III, *Xba*I, *Xho*I and *Bgl*II.

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 *Eco*RI to isolate a 1289 bp fragment, containing the UGRP promoter and exon 1, which was then ligated into an *Eco*RI-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'-CCCAAGCTTGGGTATTGGTGTCTGGAGAAGAGAG-3' and 5'-GAAGATCTGGCGGCCAGCTGCCGACCAGGCC-3'; *Hind*III and *Bgl*II cloning sites are underlined. These primers represent the putative UGRP promoter sequences from -474 to -450 and -24 to +1, relative to the translation start site, respectively (see Fig. 5). The PCR fragment generated was digested with *Hind*III and *Bgl*II and ligated into a *Hind*III- and *Bgl*II-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 htgs database for Accession number AC023855. The *Hind*III and *Bgl*II UGRP promoter fragment was then re-isolated from the pSP72 plasmid and ligated into the *Hind*III and *Bgl*II-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) respectively (Streeper *et al.* 1997, Chapman *et al.* 1999). βTC-3 cells (in 3.5 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 the lipofectamine reagent (Gibco-BRL, Gaithersburg, MD, USA) as previously described (Bischof *et al.* 2001). Following transfection, firefly and renilla luciferase activity were assayed using the Dual Luciferase Assay Kit (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'-CACCATGGAGTCCACGCTG-3' and, as a reverse primer, either 5'-TCAGGAAGAGTGGATGGGCG-3' or 5'-GGAAGAGTGGATGGGCGGT-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 × 10⁵ 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 HEPES, 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 (*M_r* 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

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, Stuke & Carman 1997) that constitutes the active site of these enzymes. The equivalent residues in human UGRP are Lys₇₂,...Arg Pro₈₀...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 six positions that differ between G6Pase catalytic subunit and IGRP; viz T₁₆>I₁₂, Q₅₄>N₅₀, A₁₂₄>S₁₂₀, L₂₁₁>K₂₀₉, S₂₉₈>C₂₉₆ and L₃₄₅>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, Q₅₄>S₅₀, G₆₈>T₆₄, T₁₀₈>S₁₀₃, G₁₂₂>I₁₁₇, A₁₂₄>G₁₁₉, W₁₅₆>L₁₄₇, G₂₂₂>L₂₁₂, T₂₅₅>I₂₄₅, N₂₆₄>D₂₅₄, L₂₆₅>S₂₅₅, R₂₉₅>K₂₈₃, S₂₉₈>C₂₈₆, V₃₃₈>L₃₂₅, 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 *M_r* (38 709 Da) (Fig. 2). The size discrepancy between the predicted and observed *M_r* 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 357aa 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 *M_r*

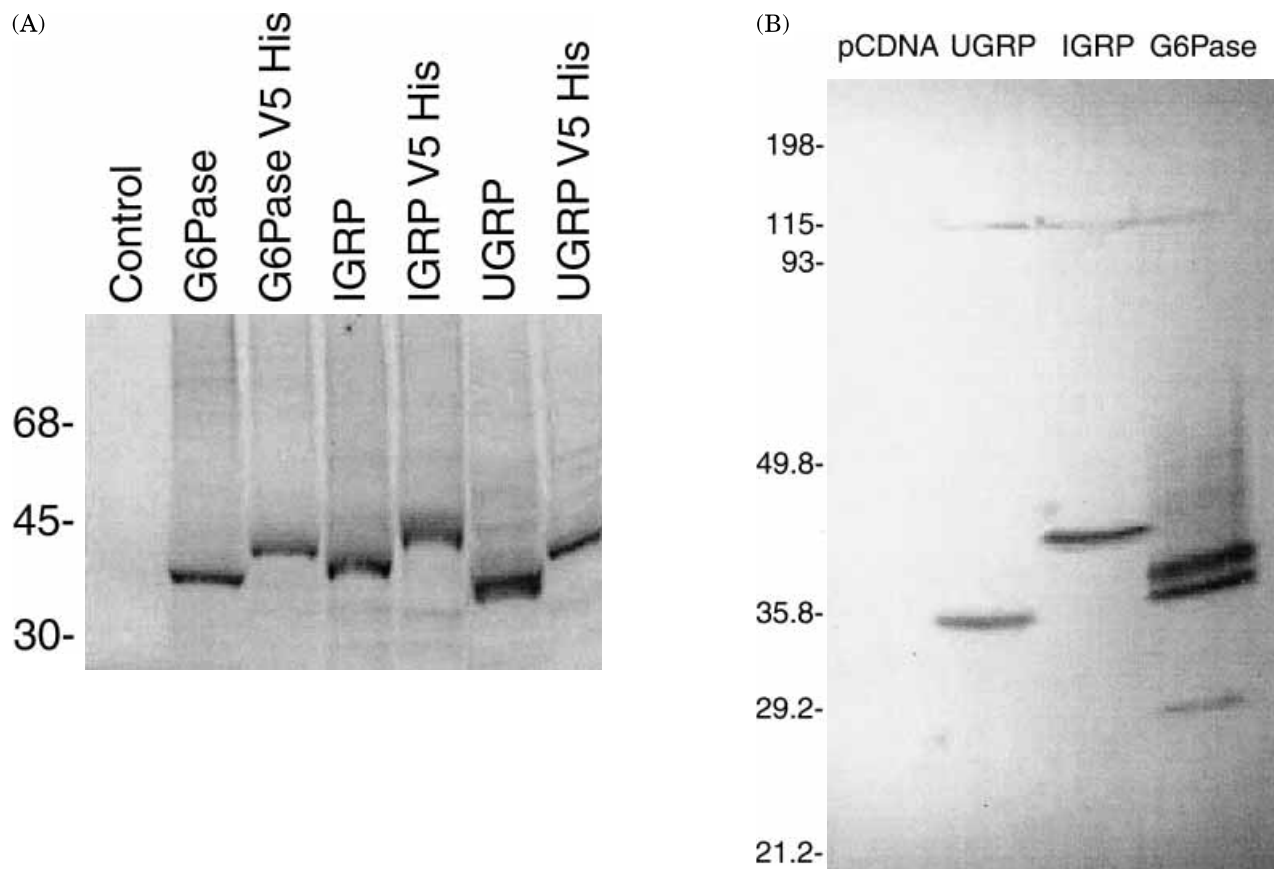


Figure 2 *In vitro* and *in situ* expression of the mouse G6Pase catalytic subunit, mouse IGRP and human UGRP. (A) T7 polymerase-derived transcripts from cloned mouse G6Pase catalytic subunit, mouse IGRP and human UGRP pcDNA3.1D constructs with or without the V5 His₆ COOH terminal extension (V5 His) were translated *in vitro* in the presence of ³⁵S 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.

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 Mg²⁺ or Ca²⁺) 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

Table 1 G6P hydrolytic activity in microsomal fractions derived from transiently transfected COS 7 cells. COS 7 cell cultures (approx. 500×10^3 cells) were co-transfected, as described in Materials and methods, with the indicated pcDNA3.1D plasmid constructs encoding human (h) UGRP, mouse (m) IGRP or the mouse G6Pase catalytic subunit (2 μ g) together with a reference vector encoding β -galactosidase (1 μ g). Enzyme activities were determined in cell homogenates after 48 h as described. Each tabulated result represents the mean \pm S.E.M. of duplicate determinations from 6 to 12 individual transfections. V5 His constructs are COOH-terminally extended versions of the molecule that incorporate a V5 epitope and His₆ tag generated by deletion of the stop codon in each insert

Sample	G6P Hydrolysis (nmol/min per mg)	β -Galactosidase (nmol/min per mg)
pcDNA3.1	2.9 \pm 0.8	528 \pm 196
hUGRP	3.7 \pm 1.1	560 \pm 255
hUGRP V5 His	4.3 \pm 1.3	526 \pm 219
mIGRP	1.5 \pm 3.3	209 \pm 66
mIGRP V5 His	1.6 \pm 0.9	581 \pm 371
mG6Pase	321.5 \pm 174.6	278 \pm 119
mG6Pase V5 His	254.0 \pm 64.0	438 \pm 94

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 \sim 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 \sim 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).

Identification and structural analysis of the UGRP gene

A BLAST search of the NCBI nr and htgs databases using the human UGRP cDNA sequence as the query identified a PAC clone (Accession number AC004150; approximately 121 kbp) and a BAC clone (Accession number AC023855; approximately 219 kbp), both of which contain the entire human UGRP 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 ORF 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.

Likewise, a BLAST search of the NCBI htgs database using the human G6Pase catalytic subunit cDNA sequence (Accession number U01120; Lei *et al.* 1993) as the query identified two BAC clones (Accession numbers AC055866 and AC046171), both of which contain the entire human G6Pase catalytic subunit gene as a contiguous sequence. The sequences of the five G6Pase catalytic subunit exons, as determined from genomic DNA, were identical to the equivalent sequences reported in the cDNA (Lei *et al.* 1993).

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 human G6Pase catalytic subunit gene exons have been previously described (Lei *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 individual PAC and BAC clones containing the UGRP gene and the two BAC clones containing the G6Pase catalytic subunit gene gave very similar estimates of intron sizes.

A comparison of these exon and intron sizes with those of the human IGRP gene (Martin *et al.* 2001) showed conservation of the exon sizes (Table 2) consistent with the three genes being evolutionarily related. The UGRP gene, however, has an additional exon (Table 2) which appears to have

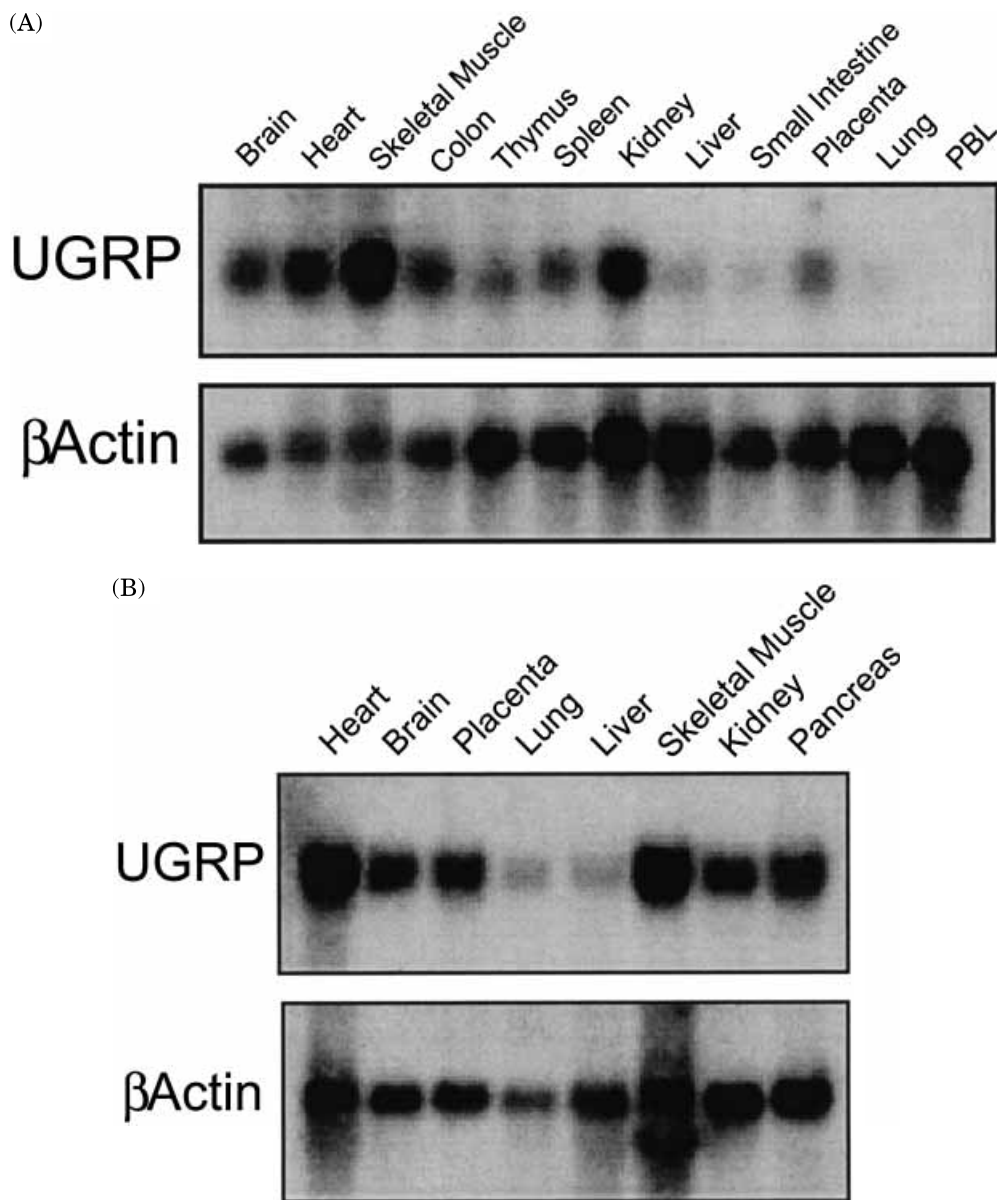


Figure 3 Analysis of human UGRP mRNA expression by RNA blotting. RNA blotting analysis was performed on the indicated ~1–2 μ g samples of poly A⁺mRNA. Clontech MTN blot #7780-1 (A) and MTN #7760-1 (B) blots were hybridized with a labeled human UGRP cDNA spanning the ORF prior to washing and visualization 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.

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

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 IGRP 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

Exon number	Exon size (bp)		
	G6Pase	IGRP	UGRP
1	309	260	218
2	110	110	107
3	106	112	91
4	116	116	119
5	>2000	~2400	142
6	N/A	N/A	662

Intron number	Intron size (bp)		
	G6Pase	IGRP	UGRP
A	2829	895	2974
B	3490	1950	413
C	1673	2046	198
D	1496	789	222
E	N/A	N/A	228

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 3 Comparison of the exon/intron boundaries of the human G6Pase catalytic subunit, IGRP and UGRP genes. The human IGRP exon/intron boundaries are from Martin *et al.* (2001) whereas the human G6Pase catalytic subunit and UGRP exon/intron boundaries were determined as described in the Results. Exon and intron sequences are shown in upper case and lower case letters respectively. The 5' and 3' consensus splice sequences are from Jackson (1991)

	Gene	5' intron junction	3' intron junction
Intron			
A	G6Pase IGRP UGRP	TTTAAGTG/gtaagaac TTTAAATG/gtaagatt TTCAAGTG/gtgagaca	ttccatag/GATTCTCT aaatacag/GATATTAT tctccag/GTTCTTT
B	G6Pase IGRP UGRP	TGGACCAG/gtaagcgt AGGTCCAG/gtaagcta TGGTCCAG/gtgggaag	cccttag/GGAGCCCC tgttgcag/GAAGTCCA gctggcag/GCAGCCCT
C	G6Pase IGRP UGRP	AGATTTCCG/gtaagaac CTGCACAG/gtcagctt GCCCCGAG/gtataacc	tgttgcag/GTGCTTGA catcgtag/ACTGACCT ctttctag/CCGCTGGG
D	G6Pase IGRP UGRP	CCTGTCAG/gtatgggc AATTGGTG/gtaaata AATAACTG/gtgagcaa	ccactcag/GCATTGCT atcctcag/GCATGCTG tctcctag/GCGCTGTC
E	G6Pase IGRP UGRP	N/A N/A CTTTCTTG/gtaagtct	N/A N/A cccacag/GTCCATCA
	Consensus sequence	(A or T)G/gtaa	cag/G 3'

N/A, not applicable.

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

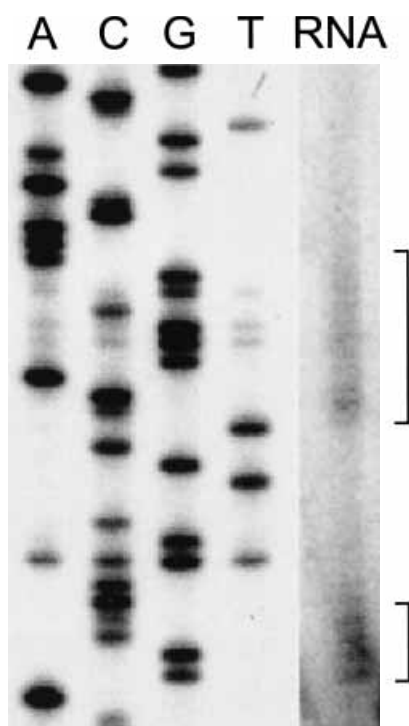


Figure 4 Determination of the human UGRP gene transcription start site. A primer extension assay was performed, as described in Materials and methods, using human muscle RNA and an oligonucleotide primer complementary to UGRP exon 1. A dideoxynucleotide sequencing reaction (labeled A, C, G and T) was also performed using the same exon 1 primer and a pGEM7 plasmid, described in Materials and methods, that contains the putative UGRP promoter and exon 1, as the template. The products of the dideoxynucleotide sequencing reaction and primer extension reaction were separated on the same denaturing polyacrylamide gel, which allowed for a direct comparison of the major extended products with their corresponding initiation base. The position of the multiple primer extension products are indicated by parentheses.

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 β TC-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; M_r 38 709) that has 36% overall identity to the human G6Pase catalytic subunit (357 aa; M_r 40 487) and 37% overall identity to human IGRP (355 aa; M_r 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

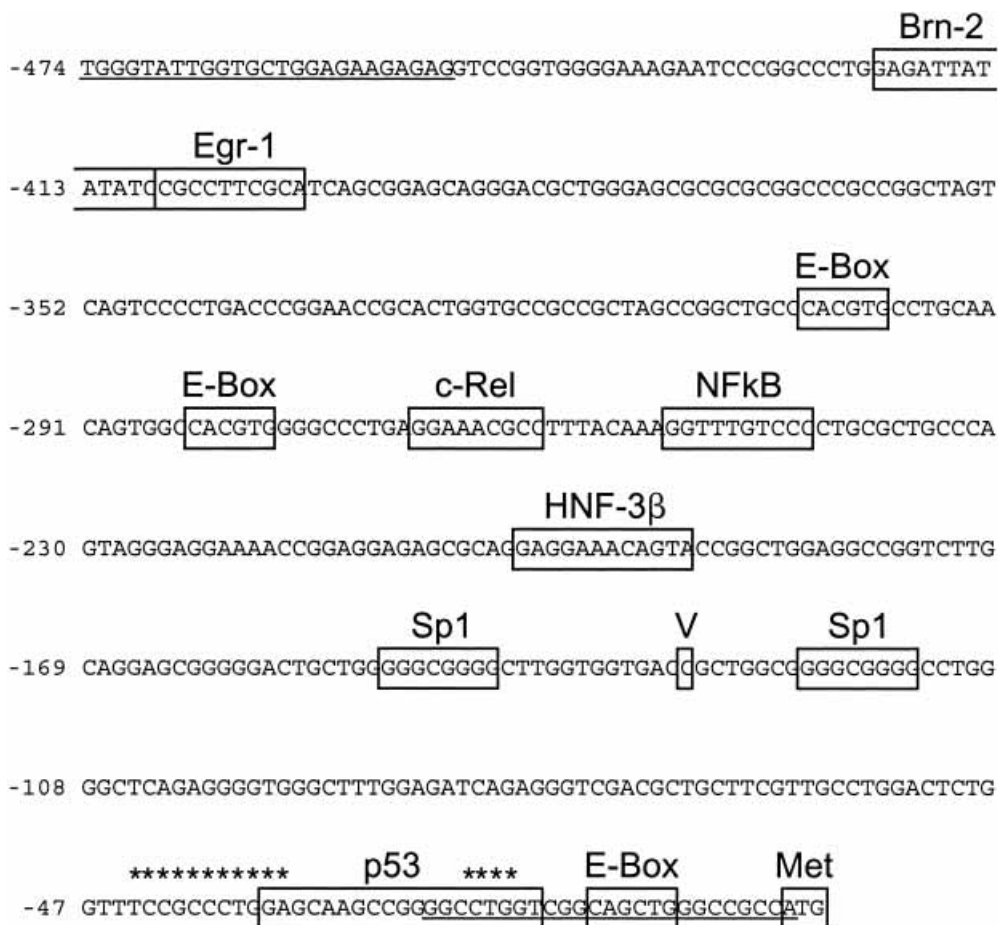
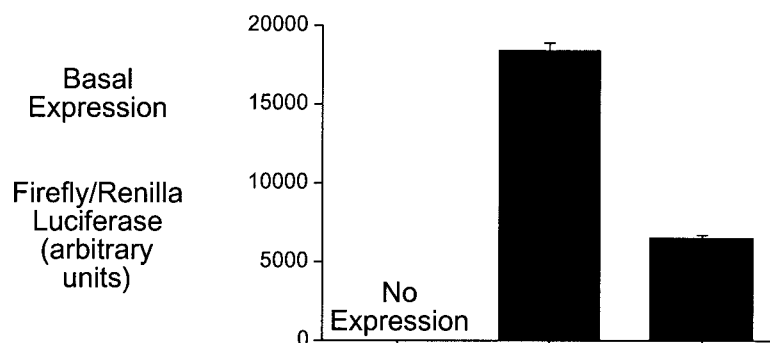


Figure 5 Sequence of the human UGRP gene promoter. The human UGRP promoter sequence is labeled relative to the predicted translation start site of UGRP, designated as +1. Putative *trans*-acting factor binding sites in the UGRP promoter, as predicted by the MatInspector sequence analysis software (Quandt *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.

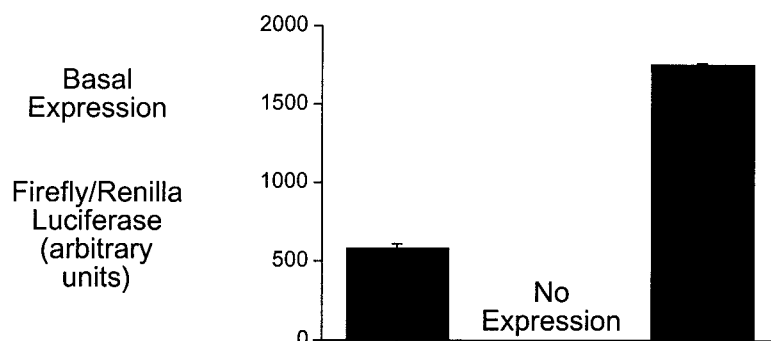
an additional coding sequence (Fig. 1 and Table 2). Major differences, however, are apparent at the level of its promoter structure; whereas the mouse IGRP and mouse G6Pase catalytic subunit promoter sequences share ~50% homology, not including spaces, over the IGRP -252 to +1 region (Ebert *et al.* 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 βTC-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

A) β TC-3 Cells

B) HepG2 Cells



C) HeLa Cells

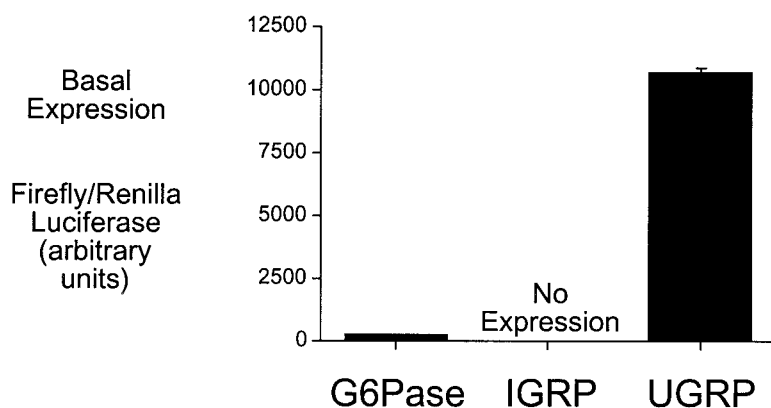


Figure 6 Comparison of the relative basal activities of the human G6Pase catalytic subunit, IGRP and UGRP gene promoters. β TC-3, HepG2 and HeLa cells were transiently co-transfected, as described in Materials and methods, with an expression vector encoding renilla luciferase and either a G6Pase catalytic 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 \pm 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 β , 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 (Streep *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 CANN TG (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 β TC-3 cell lines respectively. Research in the laboratory of R M O'Brien 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

- An J, Li Y, van De Werve G & Newgard CB 2001 Overexpression of the P46 (T1) translocase component of the glucose-6-phosphatase complex in hepatocytes impairs glycogen accumulation via hydrolysis of glucose-1-phosphate. *Journal of Biological Chemistry* **276** 10722–10729.
- Arden SD, Roep BO, Neophytou PI, Usac EF, Duinkerken G, de Vries RR & Hutton JC 1996 Imogen 38: a novel 38-kD islet mitochondrial autoantigen recognized by T cells from a newly diagnosed type 1 diabetic patient. *Journal of Clinical Investigation* **97** 551–561.
- Arden SD, Zahn T, Steegers S, Webb S, Bergman B, O'Brien RM & Hutton JC 1999 Molecular cloning of a pancreatic islet-specific glucose-6-phosphatase catalytic subunit-related protein. *Diabetes* **48** 531–542.
- Arion WJ, Canfield WK, Callaway ES, Burger HJ, Hemmerle H, Schubert G, Herling AW & Oekonomopoulos R 1998 Direct evidence for the involvement of two glucose 6-phosphate-binding sites in the glucose-6-phosphatase activity of intact liver microsomes. Characterization of T1, the microsomal glucose 6-phosphate transport protein by a direct binding assay. *Journal of Biological Chemistry* **273** 6223–6227.
- Arnold HH & Winter B 1998 Muscle differentiation: more complexity to the network of myogenic regulators. *Current Opinion in Genetic Development* **8** 539–544.
- Barnes PJ & Karin M 1997 Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases. *New England Journal of Medicine* **336** 1066–1071.
- Bischof IJ, Martin CC, Svitek CA, Stadelmaier BT, Hornbuckle LA, Goldman JK, Oeser JK, Hutton JC & O'Brien RM 2001 Characterization of the mouse islet-specific glucose-6-phosphatase catalytic subunit-related protein gene promoter by *in situ* footprinting. Correlation with fusion gene expression in the islet derived bTC-3 and hamster insulinoma tumor cell lines. *Diabetes* **50** 502–514.
- Chapman SC, Ayala JE, Streeper RS, Culbert AA, Eaton EM, Svitek CA, Goldman JK, Tavare JM & O'Brien RM 1999 Multiple promoter elements are required for the stimulatory effect of insulin on human collagenase-1 gene transcription. Selective effects on activator protein-1 expression may explain the quantitative difference in insulin and phorbol ester action. *Journal of Biological Chemistry* **274** 18625–18634.
- Cherrington AD 1999 Banting Lecture 1997. Control of glucose uptake and release by the liver *in vivo*. *Diabetes* **48** 1198–1214.
- Chou JY & Mansfield BC 1999 Molecular genetics of type I glycogen storage disease. *Trends in Endocrinology and Metabolism* **10** 104–113.
- Cline GW, Rothman DL, Magnusson I, Katz LD & Shulman GI 1994 ¹³C-nuclear magnetic resonance spectroscopy studies of hepatic glucose metabolism in normal subjects and subjects with insulin-dependent diabetes mellitus. *Journal of Clinical Investigation* **94** 2369–2376.
- Colgan DF & Manley JL 1997 Mechanism and regulation of mRNA polyadenylation. *Genes and Development* **11** 2755–2766.
- Consoli A 1992 Role of liver in pathophysiology of NIDDM. *Diabetes Care* **15** 430–441.
- Dickens M, Svitek CA, Culbert AA, O'Brien RM & Tavare JM 1998 Central role for phosphatidylinositol 3-kinase in the repression of glucose-6-phosphatase gene transcription by insulin. *Journal of Biological Chemistry* **273** 20144–20149.
- Ebert DH, Bischof IJ, Streeper RS, Chapman SC, Svitek CA, Goldman JK, Mathews CE, Leiter EH, Hutton JC & O'Brien RM 1999 Structure and promoter activity of an islet-specific glucose-6-phosphatase catalytic subunit-related gene. *Diabetes* **48** 543–551.
- Foster JD, Pederson BA & Nordlie RC 1997 Glucose-6-phosphatase structure, regulation, and function: an update. *Proceedings of the Society for Experimental Biology and Medicine* **215** 314–332.
- Gashler A & Sukhatme VP 1995 Early growth response protein 1 (Egr-1): prototype of a zinc-finger family of transcription factors. *Progress in Nucleic Acid Research and Molecular Biology* **50** 191–224.
- Gerin I, Veiga-da-Cunha M, Achouri Y, Collet JF & Van Schaftingen E 1997 Sequence of a putative glucose 6-phosphate translocase, mutated in glycogen storage disease type Ib (see comments). *FEBS Letters* **419** 235–238.
- Gerin I, Noel G & Van Schaftingen E 2001 Novel arguments in favor of the substrate-transport model of glucose-6-phosphatase. *Diabetes* **50** 1531–1538.
- Granner DK & O'Brien RM 1992 Molecular physiology and genetics of NIDDM. Importance of metabolic staging. *Diabetes Care* **15** 369–395.
- Guillemot F, Lo LC, Johnson JE, Auerbach A, Anderson DJ & Joyner AL 1993 Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* **75** 463–476.
- Haber BA, Chin S, Chuang E, Buikhuisen W, Naji A & Taub R 1995 High levels of glucose-6-phosphatase gene and protein expression reflect an adaptive response in proliferating liver and diabetes. *Journal of Clinical Investigation* **95** 832–841.
- Hara Y, Rovescalli AC, Kim Y & Nirenberg M 1992 Structure and evolution of four POU domain genes expressed in mouse brain. *PNAS* **89** 3280–3284.
- Hemrika W, Renirie R, Dekker HL, Barnett P & Wever R 1997 From phosphatases to vanadium peroxidases: a similar architecture of the active site. *PNAS* **94** 2145–2149.
- Hutton JC, Peshavaria M, Johnston CF, Ravazzola M & Orci L 1988 Immunolocalization of betagranin: a chromogranin A-related protein of the pancreatic B-cell. *Endocrinology* **122** 1014–1020.
- Jackson IJ 1991 A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Research* **19** 3795–3798.
- Jackson MR, Nilsson T & Peterson PA 1990 Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO Journal* **9** 3153–3162.
- Jacoby DB, Zilz ND & Towle HC 1989 Sequences within the 5'-flanking region of the S14 gene confer responsiveness to glucose in primary hepatocytes. *Journal of Biological Chemistry* **264** 17623–17626.
- Kornfeld R & Kornfeld S 1985 Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry* **54** 631–664.
- Kozak M 1991 Structural features in eukaryotic mRNAs that modulate the initiation of translation. *Journal of Biological Chemistry* **266** 19867–19870.
- Lai E, Clark KL, Burley SK & Darnell JE Jr 1993 Hepatocyte nuclear factor 3/fork head or 'winged helix' proteins: a family of transcription factors of diverse biologic function. *PNAS* **90** 10421–10423.
- Lassar AB, Davis RL, Wright WE, Kadesch T, Murre C, Voronova A, Baltimore D & Weintraub H 1991 Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins *in vivo*. *Cell* **66** 305–315.
- Lei KJ, Shelly LL, Pan CJ, Sidbury JB & Chou JY 1993 Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a. *Science* **262** 580–583.
- Lei KJ, Pan CJ, Shelly LL, Liu JL & Chou JY 1994 Identification of mutations in the gene for glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1a. *Journal of Clinical Investigation* **93** 1994–1999.
- Leuzzi R, Fulceri R, Marcolongo P, Banhegyi G, Zammarchi E, Stafford K, Burchell A & Benedetti A 2001 Glucose 6-phosphate transport in fibroblast microsomes from glycogen storage disease type 1b patients: evidence for multiple glucose 6-phosphate transport systems. *Biochemical Journal* **357** 557–562.

- Lin B, Morris DW & Chou JY 1997 The role of HNF1 alpha, HNF3 gamma, and cyclic AMP in glucose-6-phosphatase gene activation. *Biochemistry* **36** 14096–14106.
- Martin CC, Bischof LJ, Bergman B, Hornbuckle LA, Hilliker C, Frigeri C, Wahl D, Svitek CA, Wong R, Goldman JK, Oeser JK, Lepretre F, Froguel P, O'Brien RM & Hutton JC 2001 Cloning and characterization of the human and rat islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) genes. *Journal of Biological Chemistry* **276** 25197–25207.
- Massari ME & Murre C 2000 Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and Cellular Biology* **20** 429–440.
- Milpetz F, Argos P & Persson B 1995 TMAP: a new email and WWW service for membrane-protein structural predictions. *Trends in Biochemical Sciences* **20** 204–205.
- Mithieux G 1997 New knowledge regarding glucose-6 phosphatase gene and protein and their roles in the regulation of glucose metabolism. *European Journal of Endocrinology* **136** 137–145.
- Mithieux G 2001 New data and concepts on glutamine and glucose metabolism in the gut. *Current Opinion in Clinical Nutrition and Metabolism Care* **4** 267–271.
- Molkentin JD, Black BL, Martin JF & Olson EN 1995 Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* **83** 1125–1136.
- Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB *et al.* 1989 Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58** 537–544.
- Naya FJ, Huang HP, Qiu Y, Mutoh H, DeMayo FJ, Leiter AB & Tsai MJ 1997 Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/ neuroD-deficient mice. *Genes and Development* **11** 2323–2334.
- O'Brien RM & Granmer DK 1996 Regulation of gene expression by insulin. *Physiological Reviews* **76** 1109–1161.
- Olson EN 1990 MyoD family: a paradigm for development? *Genes and Development* **4** 1454–1461.
- Quandt K, Frech K, Karas H, Wingender E & Werner T 1995 MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Research* **23** 4878–4884.
- Sambrook J, Fritsch EF & Maniatis EF 1989 *Molecular Cloning: A Laboratory Manual*, edn 2. Plainview, NY: Cold Spring Harbor Laboratory Press.
- 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.
- Seoane J, Trinh K, O'Doherty RM, Gomez-Foix AM, Lange AJ, Newgard CB & Guinovart JJ 1997 Metabolic impact of adenovirus-mediated overexpression of the glucose-6-phosphatase catalytic subunit in hepatocytes. *Journal of Biological Chemistry* **272** 26972–26977.
- Smale ST 1997 Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* **1351** 73–88.
- Streeper RS, Svitek CA, Chapman S, Greenbaum LE, Taub R & O'Brien RM 1997 A multicomponent insulin response sequence mediates a strong repression of mouse glucose-6-phosphatase gene transcription by insulin. *Journal of Biological Chemistry* **272** 11698–11701.
- Streeper RS, Eaton EM, Ebert DH, Chapman SC, Svitek CA & O'Brien RM 1998 Hepatocyte nuclear factor-1 acts as an accessory factor to enhance the inhibitory action of insulin on mouse glucose-6-phosphatase gene transcription. *PNAS* **95** 9208–9213.
- Streeper RS, Svitek CA, Goldman JK & O'Brien RM 2000 Differential role of hepatocyte nuclear factor-1 in the regulation of glucose-6-phosphatase catalytic subunit gene transcription by cAMP in liver- and kidney-derived cell lines. *Journal of Biological Chemistry* **275** 12108–12118.
- Streeper RS, Hornbuckle LA, Svitek CA, Goldman JK, Oeser JK & O'Brien RM 2001 Protein kinase A phosphorylates hepatocyte nuclear factor-6 and stimulates glucose-6-phosphatase catalytic subunit gene transcription. *Journal of Biological Chemistry* **276** 19111–19118.
- Struhl K 1998 Histone acetylation and transcriptional regulatory mechanisms. *Genes and Development* **12** 599–606.
- Stukey J & Carman GM 1997 Identification of a novel phosphatase sequence motif. *Protein Science* **6** 469–472.
- Stumvoll M, Meyer C, Mitrakou A, Nadkarni V & Gerich JE 1997 Renal glucose production and utilization: new aspects in humans. *Diabetologia* **40** 749–757.
- Suske G 1999 The Sp-family of transcription factors. *Gene* **238** 291–300.
- Trinh KY, O'Doherty RM, Anderson P, Lange AJ & Newgard CB 1998 Perturbation of fuel homeostasis caused by overexpression of the glucose-6-phosphatase catalytic subunit in liver of normal rats. *Journal of Biological Chemistry* **273** 31615–31620.
- Van Schaftingen E & Gerin I 2002 The glucose-6-phosphatase system. *Biochemical Journal* **362** 513–532.
- Veiga-da-Cunha M, Gerin I, Chen YT, de Barsey T, de Lonlay P, Dionisi-Vici C, Fenske CD, Lee PJ, Leonard JV, Maire I, McConkie-Rosell A, Schweitzer S, Vikkula M & Van Schaftingen E 1998 A gene on chromosome 11q23 coding for a putative glucose-6-phosphate translocase is mutated in glycogen-storage disease types Ib and Ic. *American Journal of Human Genetics* **63** 976–983.
- Vogelstein B, Lane D & Levine AJ 2000 Surfing the p53 network. *Nature* **408** 307–310.
- van de Werve G, Lange A, Newgard C, Mechin MC, Li Y & Berteloot A 2000 New lessons in the regulation of glucose metabolism taught by the glucose 6-phosphatase system. *European Journal of Biochemistry* **267** 1533–1549.

Received in final form 12 June 2002

Accepted 19 June 2002