Two Homologues Encoding Human UDP-Glucose:Glycoprotein Glucosyltransferase Differ in mRNA Expression and Enzymatic Activity†

Stacey M. Arnold,‡ Liselotte I. Fessler,§ John H. Fessler,§ and Randal J. Kaufman*,‡,||

Department of Biological Chemistry and the Howard Hughes Medical Institute, University of Michigan Medical Center, Ann Arbor, Michigan 48109, and Molecular Biology Institute and MCD-Biology Department, University of California, Los Angeles, California 90095-1570

Received July 19, 1999; Revised Manuscript Received November 2, 1999

ABSTRACT: UDP-glucose:glycoprotein glucosyltransferase (UGT) is a soluble protein of the endoplasmic reticulum (ER) that operates as a gatekeeper for quality control by preventing transport of improperly folded glycoproteins out of the ER. We report the isolation of two cDNAs encoding human UDP-glucose: glycoprotein glucosyltransferase homologues. HUGT1 encodes a 1555 amino acid polypeptide that, upon cleavage of an N-terminal signal peptide, is predicted to produce a soluble 173 kDa protein with the ER retrieval signal REEL. HUGT2 encodes a 1516 amino acid polypeptide that also contains a signal peptide and the ER retrieval signal HDEL. HUGT1 shares 55% identity with HUGT2 and 31−45% identity with *Drosophila*, *Caemorhabditis elegans*, and *Schizosaccharomyces pombe* homologues, with most extensive conservation of residues in the carboxy-terminal fifth of the protein, the proposed catalytic domain. HUGT1 is expressed as multiple mRNA species that are induced to similar extents upon disruption of protein folding in the ER. In contrast, HUGT2 is transcribed as a single mRNA species that is not induced under similar conditions. HUGT1 and HUGT2 mRNAs are broadly expressed in multiple tissues and differ slightly in their tissue distribution. The HUGT1 and HUGT2 cDNAs were expressed by transient transfection in COS-1 monkey cells to obtain similar levels of protein localized to the ER. Extracts from HUGT1-transfected cells displayed a 27-fold increase in the transfer of [^{14}C]glucose from UDP-[^{14}C]-glucose to denatured substrates. Despite its high degree of sequence identity with HUGT1, the expressed recombinant HUGT2 protein was not functional under the conditions optimized for HUGT1. Site-directed alanine mutagenesis within a highly conserved region of HUGT1 identified four residues that are essential for catalytic function.

In eukaryotes, transmembrane proteins and soluble proteins destined for secretion or organelle localization are cotranslationally translocated into the lumen of the endoplasmic reticulum (ER).[1,2] Within this compartment, protein folding and assembly of multimeric protein complexes are dictated by the primary amino acid sequence and by posttranslational modifications. These folding reactions are facilitated by transient association with molecular chaperones (*1*). For the majority of glycoproteins, the rate-limiting step in delivery to organelles or to the cell surface is export from the ER to the Golgi apparatus (*2*). Extensive studies have shown that the most important factor that contributes to the rate of transport from the ER to the Golgi is the rate by which proteins attain their native folded conformation. Processes designed to ensure selective transport of properly folded functional proteins rely upon mechanisms that prevent the transport of unfolded, misfolded, or unassembled proteins out of the ER and have been termed "quality control" (*3*). Recent studies have indicated that one process fundamental to the retention of glycoproteins in the ER is the processing of asparagine(N)-linked oligosaccharides. In this report, we describe the expression and activity of two human cDNAs that encode homologues of a pivotal component in the regulation of this quality control, the UDP-Glc:glycoprotein glucosyltransferase (UGT).

High-mannose-containing oligosaccharide core precursors having the structure $Glc_3Man_9GlcNAc_2$ are attached cotranslationally to asparagine residues of the consensus recognition site Asn-Xxx-Ser/Thr within nascent polypeptides translocated into the ER (*4*). Subsequently, the terminal α(1,2)-linked glucose is removed by glucosidase I, a membrane-associated homotetramer (*5*), and the two α(1,3)-linked glucose residues are removed by glucosidase II, a heterodimer peripherally associated with the ER membrane (*6*,

† Portions of this work were supported by NIH Grant AG-02128/GM-57689 (LIF and JHF) and PO1HL57345 and HL53777 (RJK) and NIH Grant GM07767 (SMA).

‡ Department of Biological Chemistry, University of Michigan Medical Center.

§ Molecular Biology Institute, University of California.

|| Howard Hughes Medical Institute, University of Michigan Medical Center.

* To whom correspondence should be addressed. Telephone: 734-763-9037. Fax: 734-763-9323. E-mail: kaufmanr@umich.edu.

[1] The abbreviations used are: HUGT, human UDP-glucose:glycoprotein glucosyltransferase; UGT, UDP-glucose:glycoprotein glucosyltransferase; EST, expressed sequence tag; ER, endoplasmic reticulum; Glc, glucose; Man, mannose; GlcNac, *N*-acetylglucosamine; PCR, polymerase chain reaction; PBS, phosphate-buffered saline; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TCA, trichloroacetic acid; EDTA, ethylenediaminetetraacetic acid; wt, wild-type; FITC, fluorescein isothiocyanate; DMEM, Dulbecco's modified essential medium; DEAE-dextran, diethylaminoethyl dextran.

[2] The sequences for HUGT1 and HUGT2 have been submitted to the GenBank and have been assigned the accession numbers AF227905 and AF227906, respectively.

10.1021/bi9916473 CCC: $19.00
Published on Web 02/09/2000

7). In addition, removal of one or two α(1,2)-linked peripheral mannose residues occurs in the lumen of the ER. The glucose-free, high-mannose-containing oligosaccharides are then subject to re-glucosylation, resulting in the formation of $Glc_1Man_{7-9}GlcNAc_2$ structures. The majority of glycoproteins are potential substrates for this re-glucosylation (*8*). These $Glc_1Man_{7-9}GlcNAc_2$ structures are then immediately deglucosylated in vivo. The enzymes that catalyze the cycle of transient re- and de-glucosylation are the UDP-Glc: glycoprotein glucosyltransferase and glucosidase II, respectively (*9*).

Recent studies have identified protein chaperones that interact with glycoprotein intermediates in the glucose trimming process. Two ubiquitously expressed, lectin protein chaperones, calnexin and calreticulin, interact with monoglucosylated forms of these N-linked glycans to prevent exit of unfolded glycoproteins from the ER (*10*−*15*). Calnexin is a 90 kDa, highly conserved, type-1 integral membrane ER protein (*16*). Calreticulin, a 46 kDa ER lumenal polypeptide, is the major Ca^{2+}-binding protein of the cell (*17*). These two chaperones, though 42−78% identical, bind to an overlapping nonidentical group of glycoproteins (*18*). The interactions are reported to promote proper folding, prevent premature oligomerization, and inhibit degradation of a variety of glycoproteins [see ref *19* for review]. Exposure of the calnexin or calreticulin ligand, the monoglucosylated core N-linked oligosaccharide, may result from the initial glucose trimming process or from transient re-glucosylation. Release from calnexin or calreticulin occurs coincidentally with removal of the glucose residue by glucosidase II and just prior to the transit of completely folded glycoproteins to the Golgi compartment (*20*). Incompletely folded, misfolded, or unassembled proteins are specific targets for re-glucosylation by UDP-Glc:glycoprotein glucosyltransferase (UGT) (*20*, *21*), a soluble resident protein of the ER. This addition of a single glucose residue allows a glycoprotein to engage in another round of chaperone interaction, and the cycle continues until the protein is properly folded (*20*), or until a protein incapable of folding is shuttled off to a degradative pathway (*22*). In this model, UGT provides a pivotal function through its ability to selectively re-glucosylate unfolded glycoproteins so that they are retained within the ER (*19*).

UGT activities from *Trypanosoma cruzi*, *S. pombe*, *Drosophila*, and rat have been partially characterized. The *Drosophila* protein was isolated and localized by immunoelectron microscopy to the ER and to the intermediate compartment leading to the Golgi apparatus (*23*). The rat liver enzyme has apparent molecular weights of 150 and 270 kDa under denaturing and native conditions, respectively (*9*), a finding consistent with UGT dimerization. UGT has an enzymatic preference for a neutral pH and for divalent cation cofactors including Ca^{2+} and Mn^{2+} but not Mg^{2+}. UGT specifically recognizes two elements in substrate glycoproteins: the innermost *N*-acetylglucosamine unit of the N-linked oligosaccharide core and patches of hydrophobic residues exposed on proteins in their non-native conformation (*21*, *24*). These elements must be present on the same protein and both are required for the transfer of glucose from the precursor UDP-glucose. Since misfolded, partially folded, or unassembled proteins are substrates for re-glucosylation by UGT, only properly folded glycoproteins may exit the ER. Therefore, UGT functions as a central regulator of quality control for protein transport out of the ER (*15*, *25*). Although the rat UGT has been purified and characterized, sequence information for a mammalian homologue has not been reported. To isolate additional modulators of glycoprotein folding and transport, we have identified two cDNAs encoding human homologues of UGT and characterized their expression and activity in mammalian cells.

MATERIALS AND METHODS

Isolation of cDNAs Encoding Homologues of UGT. The following primers were used in a polymerase chain reaction (PCR) to amplify a 438 bp sequence from *Drosophila* UDP-Glc:glycoprotein glucosyltransferase (DUGT) cDNA (*23*): 5′-TTCCAGTACGAATTGGTCCAG-3′ (sense primer corresponding to coding nucleotides 4089−4109) and 5′-GTTGGGCAAGTCCTGATCCAA-3′ (complement primer corresponding to coding nucleotides 4506−4526). The amplified fragment was radiolabeled by random-priming (Rediprime Kit and [α^{32}P]-dCTP, Amersham Pharmacia Biotech, Inc., Piscataway, NJ) and used as a probe to screen a human fetal liver cDNA library in λgt10 (Clontech, Palo Alto, CA). Positive phage were isolated and two overlapping cDNA inserts (Figure 1A, Clones 1 and 2) were subcloned into the *Eco*RI restriction site of Bluescript II KS (Stratagene, La Jolla, CA) for DNA sequencing. The 2.6 kb cDNA clone was labeled by random priming and used to screen the λgt10 library for the additional 5′ sequence. An approximately 2.8 kb cDNA clone (Figure 1A, Clone 3) was obtained that shared 742 bp of identical sequence overlap with Clone 1 and extended the 5′ open reading frame by approximately 1.7 kb. It also contained, at its 5′-end, 342 bp of what appeared to be an intron sequence, with termination codons in all reading frames. Therefore, a complement primer to the 5′-most in-frame sequence sharing homology with the *Drosophila* cDNA (5′-TATCATGCAATATTGGAGGCT-3′, corresponding to coding nucleotides 295−315) was then used in conjunction with one λgt10 vector-specific primer to obtain the remaining 282 bp of 5′-coding sequence information by PCR amplification from the cDNA library. The longest PCR products contained a potential initiation codon in a favorable context for initiation (*26*) as well as 38 bp of upstream sequence that lacked an ATG codon. The remaining 3′ sequence (extending up to and including the poly $(A)^+$ tail) was obtained by reverse transcription and PCR of total RNA isolated from HeLa cells.

A second homologous sequence was identified by search of the Genbank human expressed sequence tags (EST) database with the sequence information obtained above. The search identified homologous but nonidentical ESTs with the following accession numbers: AA225611, AA225666, AA400517, AA424587, AA424632, AA515912, AA620797, AA633581, AA740607, AA757270, H00748, HSC2MH052, N32107, N56779, R71439, W37208, and W51817. These ESTs were compiled in an overlapping fashion using the GCG programs PILEUP and PRETTY. To obtain the full coding region, the insert from EST AA515912, comprising the 5′-most sequence information (coding nucleotides 3240−3611), was excised and used as a probe to screen the same fetal liver cDNA library. A 2.6 kb fragment was isolated by PCR amplification using the λgt10 vector primers and subcloned into the PCR 2.1 vector (Invitrogen, Carlsbad,

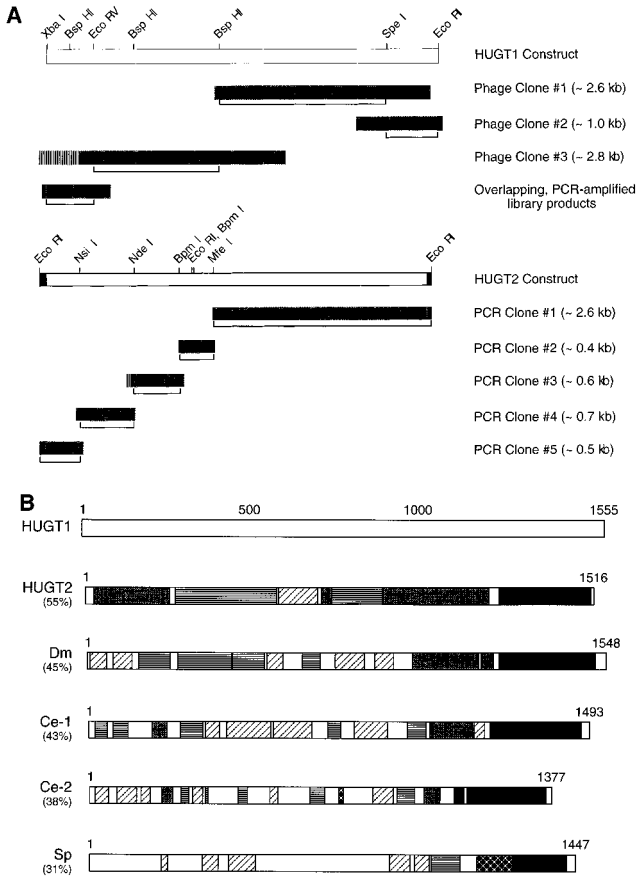


FIGURE 1: Assembly of HUGT cDNAs and sequence alignments. Panel A: Phage clone and products amplified by PCR from the λgt10 library are aligned with their final positions in the full-length HUGT1 and HUGT2 constructs. Shaded regions indicate the following: (top bar) full-length coding region; (grey shade) clone sequence contained within the coding region; (vertical stripes) probable intron sequence; (black shade) sequence representing the untranslated region. Brackets underneath clones indicate portions incorporated into the final sequences. Restriction sites used for subcloning are indicated. Panel B: alignment of HUGT1 with HUGT2 and with other known and putative UGT sequences from *Drosophila melanogaster* (Dm), *C. elegans* (Ce-1 and Ce-2, accession numbers 860712 and CAB03874, respectively), and *S. pombe* (Sp).. HUGT1 and HUGT2 were aligned by BLAST comparison of the two sequences. The GCG GAP and NCBI BLAST programs were used to align the UGTs from other species. Areas of identity are shaded according to the following scheme: (diagonal lines) 20−39% identity; (horizontal lines) 40−49% identity; (dark grey shade) 50−59% identity; (cross pattern) 60−69% identity; (black shade) 70−79% identity. Overall identity with HUGT1 is indicated in parentheses to the left of each homologue.

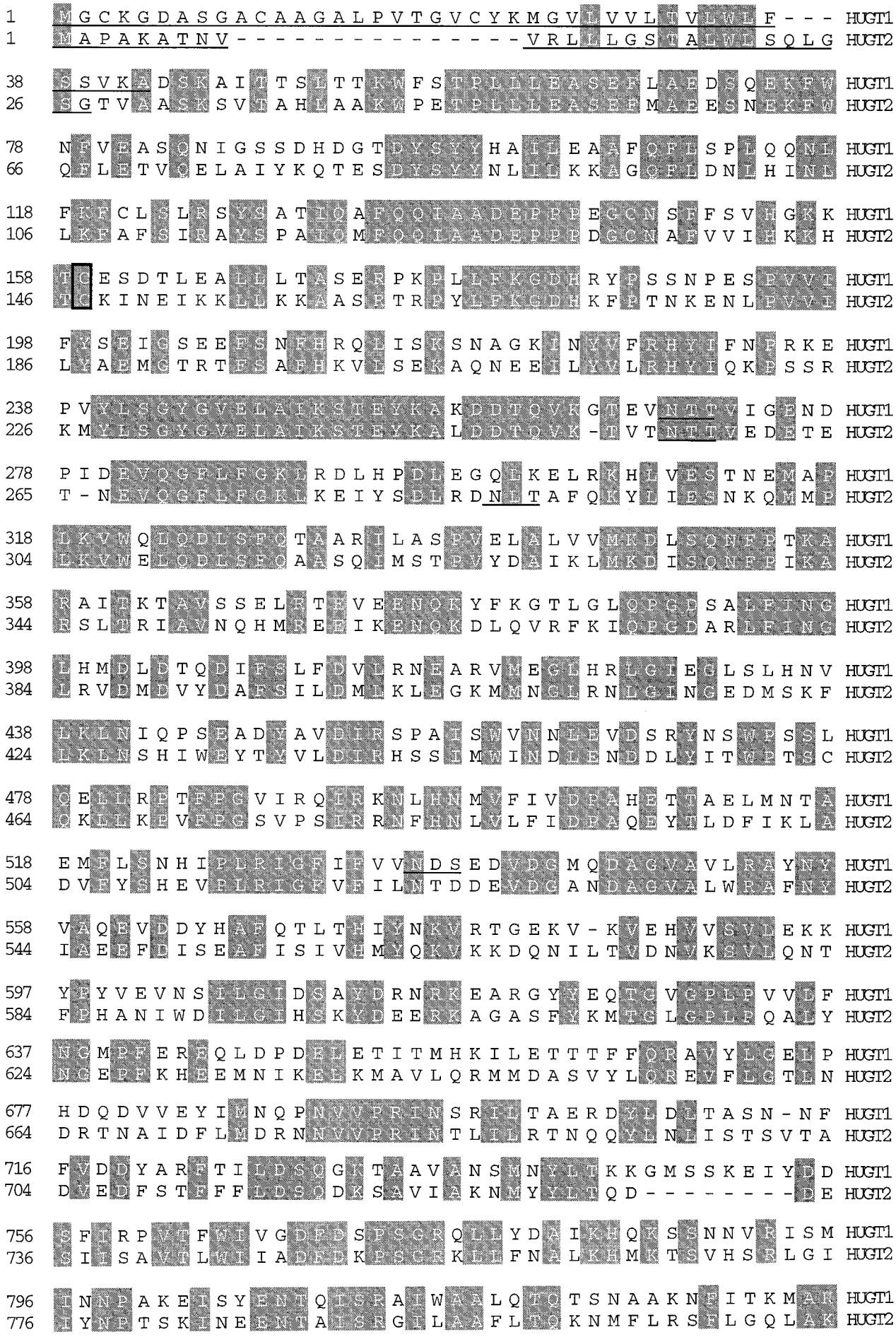
CA). The remaining sequence information (see Figure 2) was obtained by stepwise PCR amplification using primers specific to the HUGT2 sequence in combination with the λgt10 vector primers. Primary PCR amplification at each step was followed by amplification with nested primers to ensure sequence specificity of the products being produced. The resulting amplification products were subcloned into the PCR 2.1 vector and sequenced. Clones 1 and 2 (Figure 1A) shared 26 bp of identical overlap. Clones 2 and 3 (Figure 1A) shared 67 bp of identical overlap. Clones 3 and 4 (Figure 1A) shared 39 bp of identical overlap. Clones 4 and 5 (Figure 1A) shared 83 bp of identical overlap. Clone 5 (Figure 1A) contained a potential initiation codon in a favorable context for initiation (*26*) as well as 71 bp of upstream sequence that lacked an ATG codon. The location of the polyadenylation signal was determined by analysis of the overlapping EST clones listed above.

UGT mRNA Expression. Fragments underlined in Figure 3A were amplified by PCR and subcloned into Bluescript II KS. Vector inserts were radiolabeled by random-priming and used to probe a multiple tissue Northern blot (Clontech). ExpressHyb Hybridization Solution (Clontech) was used, according to the suggested protocol. The 5′-most portion of HUGT1 Clone 3, encoding 1 kb of the coding region, was excised from the corresponding λgt10 vector via *Stu*1 and *Eco*R1 restriction endonuclease digestion, radiolabeled by random-priming, and used to probe the same multiple human tissue Northern blot via the ExpressHyb protocol.

Exponentially growing COS-1 cells were treated with 10 μg/mL tunicamycin or 5 μM A23187 for the indicated periods of time. At each time point, total RNA was collected in Trizol reagent (Gibco BRL, Gaithersburg MD). Poly $(A)^+$ RNA was isolated by chromatography on an oligo-dT column (Gibco BRL), denatured with dimethyl sulfoxide/glyoxal, and electrophoresed on an agarose gel (*27*). The RNA was transferred in a solution of 7.5 μM NaOH to a Hybond N+ positively charged nylon membrane (Amersham Pharmacia Biotech, Inc.). The membrane was hybridized with the following probes prepared by random-priming; the 5′ 1 kb HUGT1 fragment described above, the HUGT1 and HUGT2 3′ fragments underlined in Figure 3A, a 1.5 kb *Eco*RI fragment from hamster BiP cDNA (*28*), and a 900 bp *Not*1/*Eco*RV fragment from chicken β-actin cDNA (*29*). Band intensities were measured by PhosphorImager analysis and quantified by comparison to β-actin control.

Expression of HUGTs in COS-1 Cells. For assembly of the full-length HUGT1 sequence, an *Spe*1/*Eco*R1 fragment from Clone 2 (Figure 1A) was used to replace the corresponding fragment of Clone 1 in the Bluescript II KS vector. This combined sequence, Clone 1-2, was subsequently excised from the *Eco*R1 site of Bluescript and subcloned into the *Eco*R1 site of pEDΔ*C* (*30*). An *Eco*R1/*Bsp*H1 fragment generated from Clone 3 replaced the corresponding fragment in the Clone 1-2 pEDΔ*C* vector; the presence of multiple *Bsp*H1 sites in Clone 3 (Figure 1A) and the presence of three additional sites in the vector required the use of partial *Bsp*H1 endonuclease digestion for the generation of the correct insert and vector fragments. Primers, 5′-GC*T-CTAGA*CCACC**ATG**GGCTGCAAGGGAGAC-3′ (sense primer with sequence corresponding to coding nucleotides 1−18 and including an initiation codon, in bold, in a favorable context for translation (*26*) as well as an *Xba*1 site, in italics) and 5′-GCCTTGTACTCAGTGCTCTTA-3′ (complement primer corresponding to coding nucleotides 750−770) were designed to PCR-amplify the remaining 5′-sequence, and the resulting fragment was subcloned, via the *Xba*1 and *Eco*RV restriction sites, into the expression vector described above. The resulting expression plasmid DNA was designated HUGT1 and the cDNA was sequenced in its entirety.

For assembly of full-length HUGT2, an *Nsi*1 fragment from Clone 5 (one site present in overlapping region of Clones 4 and 5, and site in the vector) was generated to replace the corresponding portion of Clone 4. A partial digest was performed with *Bpm*1 on Clone 2 (one site present in overlapping region of Clones 2 and 3, one additional site in


1 MGCKGDASGACAAGALPVTGVCYKMGVLVVLTVLWLF--- HUGT1
1 MAPAKATNV----------VRLLLGSTALWLSQLG HUGT2
38 SSVKADSKAITTSLTTKWFSTPLLEASEFLAEDSQEKFW HUGT1
26 SGTVAASKSVTAHLAAKWPETPLLEASEPMAESNEKFW HUGT2
78 NFVEASQNIGSSDHDGTDYSYYHAILEAAFQFLSPLQQNL HUGT1
66 QFLETVQELAIYKQTESDYSYYNLILKKAGQFLDNIHINL HUGT2
118 FKFCLSLRSYSATIQAFQQIAADEPPPEGCNSFFSVHGKK HUGT1
106 LKFAFSIRAYSPAIQMFQQLAADEPPPDGCNAFVVIHKKH HUGT2
158 TCESDTLEALLLTASERPKPLLFKGDHRYPSSNPESFVVI HUGT1
146 TCKINEIKKLLKKAASRTRPYLFKGDHKFPTNKENLFVVI HUGT2
198 FYSEIGSEEFSNFHRQLISKSNAGKINYVFRHYIFNFRKE HUGT1
186 LYAEMGTRTFSAFHKVLSEKAQNEEILYVLRHYIQKPSSR HUGT2
238 PVYLSGYGVELAIKSTEYKAKDDTQVKGTEVNTTVIGENDHUGT1
226 KMYLSGYGVELAIKSTEYKALDDTQVK-TVTNTTVEDETE HUGT2
278 PIDEVQGFLFGKLRDLHPDLEGQLKELRKHLVESTNEMAF HUGT1
265 T-NEVQGFLFGKLKEIYSDLRDNLTAFQKYLIESNKQMMP HUGT2
318 LKVWQLQDLSFQTAARILASPVELALVVMKDLSQNFPTKA HUGT1
304 LKVWELQDLSEQAASQIMSTPVYDAIKLMKDISQNFPIKA HUGT2
358 RAITKTAVSSELRTEVEENQKYFKGTLGLQPGDSALFING HUGT1
344 RSLTRIAVNQHMREEIKENQKDLQVRFKIQPGDARLFING HUGT2
398 LHMDLDTQDIFSLFDVLRNEARVMEGLHRLGIEGLSLHNV HUGT1
384 LRVDMDVYDAFSILDMLKLEGKMMNGLRNLGINGEDMSKF HUGT2
438 LKLNIQPSEADYAVDIRSPAISWVNNLEVDSRYNSWPSSL HUGT1
424 LKLNSHIWEYTYVLDIRHSSTMWINDLENDDLYITWPTSC HUGT2
478 QELLRPTFPGVIRQIRKNLHNMVFIVDPAHETTAELMNTA HUGT1
464 QKLLKPVFPGSVPSIRRNFHNLVLFIDPAQEYTLDFIKLA HUGT2
518 EMFLSNHIPLRIGFIFVVNDSEDVDGMQDAGVAVLRAYNY HUGT1
504 DVFYSHEVPIRIGFVFILNTDDEVDGANDAGVALWRAFNY HUGT2
558 VAQEVDDYHAFQTLTHIYNKVRTGEKV-KVEHVVSVLEKK HUGT1
544 IAEEFDISEAFISIVHMYQKVKKDQNILTVDNVKSVLQNT HUGT2
597 YPYVEVNSILGIDSAYDRNRKEARGYYEQTGVGPLPVVLF HUGT1
584 FPHANIWDILGIHSKYDEERKAGASFYKMTGLGPLPQALY HUGT2
637 NGMPFERFQLDPDFLETITMHKILETTTFFQRAYYLGELP HUGT1
624 NGEPFKHEEMNIKELKMAVLQRMMDASVYLQREVFLGTLN HUGT2
677 HDQDVVEYIMNQPNVVPRINSRILTAERDYLDLTASN-NF HUGT1
664 DRTNAIDFLMDRNNVVPRINTLTLRTNQQYLNLISTSVTA HUGT2
716 FVDDYARFTILDSQGKTAAVANSMNYLTKKGMSSKEIYDD HUGT1
704 DVEDFSTFFFLDSQDKSAVIAKNMYYLTQD--------DE HUGT2
756 SFIRPVTFWIVGDFDSPSGRQLLYDAIKHQKSSNNVRISM HUGT1
736 SIISAVTLWIIADFDKPSCEKLLFNALKHMKTSVHSRLGI HUGT2
796 INNPAKEISYENTQISRAIWAALQTQTSNAAKNFITKMAR HUGT1
776 IYNPTSKINEENTAISRGILAAFLTQKNMFLRSFLGQLAK HUGT2

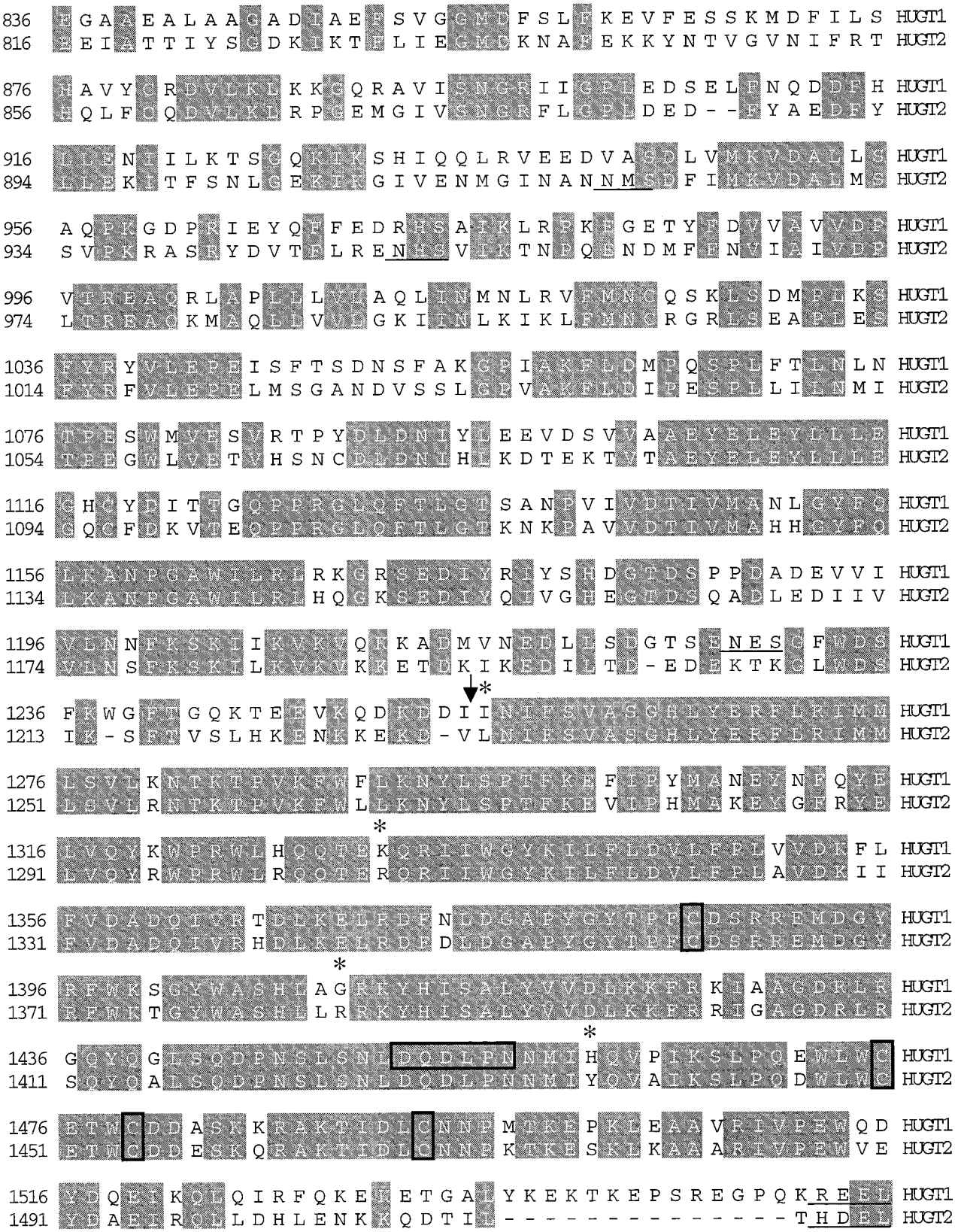


FIGURE 2: Comparison of HUGT1 and HUGT2 protein sequences, with identical residues enclosed in shaded boxes. The predicted signal peptide determined by the SignalP program (*45*), potential N-glycosylation sites, and carboxy-terminal ER retrieval signals are underlined. The cysteine residues conserved in putative UDP-Glc:glycoprotein glucosyltransferases and the six consecutive amino acids deleted in ΔHUGT are enclosed in a box (□). The boundary of the proposed catalytic domain is identified (arrow). Catalytic domain residues conserved in all homologues but HUGT2 are denoted with an asterisk.

Clone 2, one vector site) to obtain a fragment for replacement of the corresponding region of Clone 3 to obtain Clone 2-3. An *Nde*1/*Xho*1 (*Nde*1 site in region of overlap between Clone 2-3 and Clone 4-5; *Xho*1 site in vector) fragment of Clone 4-5 replaced the corresponding region of Clone 2-3 to yield Clone 2−5. An *Mfe*1/*Bam*H1 (*Mfe*1 site in region of overlap

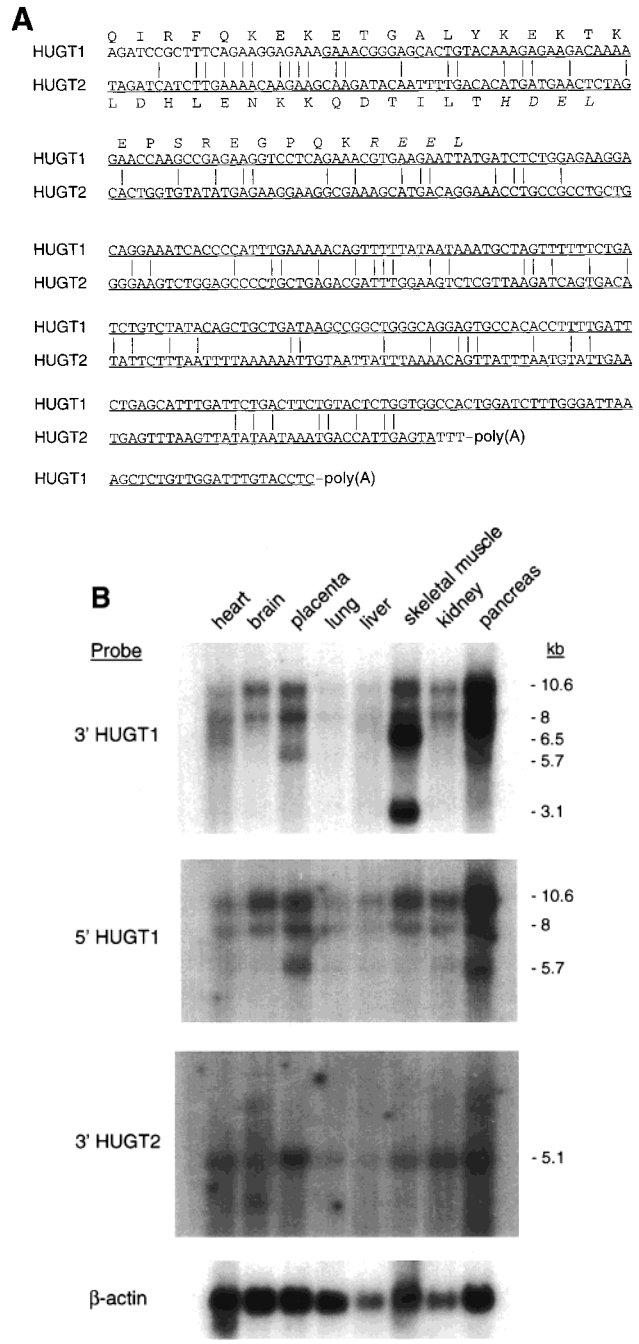


FIGURE 3: HUGT1 and HUGT2 display differential tissue expression. Panel A: The 3′-regions used to distinguish between HUGT1 and HUGT2 mRNAs are aligned, with the translation products for the portions of coding region indicated above (HUGT1) or below (HUGT2). Retrieval signals are in italics. Underlined nucleic acid sequences represent probes used for the multiple tissue Northern blot and for the analysis of mRNA inducibility. Panel B: A multiple human tissue Northern blot containing poly $(A)^+$ RNA was probed to determine the expression patterns of HUGT1 and HUGT2 mRNAs. Shown from top to bottom are the hybridizations to a 3′ HUGT1 probe, a 1 kb 5′-HUGT1 probe, a 3′-HUGT2 probe, and a β-actin probe to control for differences in RNA loading. Molecular weights of the major mRNA species are shown on the right.

between Clone 2-5 and Clone 1 and in additional region of Clone 1; *Bam*H1 site in vector) fragment of Clone 1, produced by partial digestion, replaced the corresponding region of Clone 2-5. The full-length product was subsequently excised from PCR 2.1 via *Eco*R1 digestion and inserted into the *Eco*R1 site of pEDΔC. The assembled HUGT2 expression construct was verified by DNA sequence analysis.

A deletion of six consecutive HUGT1 amino acid residues, amino acids 1452−1457, was generated by overlap-extension PCR mutagenesis. Primers *5′-CGAGTAATAACTTCTTTGTGGA-3′ (sense primer corresponding to coding nucleotides 2132−2153) and 5′-ATGAATCATGTTAAGATTTGAAAG-3′ (complement primer corresponding to coding nucleotides 4342−4353 and 4372−4383) were used to generate the 5′ fragment and primers 5′-CTTTCAAATCTTAACATGATTCAT-3′ (sense primer corresponding to coding nucleotides 4342−4353 and 4372−4383) and *5′-G*GAATTC*CGGAGACAGATCA-3′ (complement primer corresponding to the 3′ untranslated region, *Eco*R1 site in italics) were used to generate the 3′ fragment. Primers designated by asterisks were then used to amplify the overlapping fragments for substitution, via *Spe*1 and *Eco*R1 sites, into the HUGT1 expression vector described above. The mutation was confirmed by DNA sequencing, and the resulting plasmid DNA was designated ΔHUGT1. Wild-type and mutant plasmids were prepared by centrifugation in cesium chloride density gradients (*27*) and verified by restriction endonuclease digestion. Point mutants were generated in an analogous fashion, with the following exceptions: the 5′-flanking primer described above (corresponding to nucleotides 2132−2153) was replaced by 5′-GGGTTACAAGATCCTCTTCC-3′ (corresponding to nucleotides 4008−4027), and each sense overlapping primer replaced a triplet (codon) in the sequence 5′-GAT CAA GAT CTG CCC AAT-3′ (coding nucleotides 4343−4352, translated DQDLPN) with GCC (CGG in each corresponding complement overlapping primer) to code for alanine. Resulting plasmids were designated D1452A, Q1453A, D1454A, L1455A, P1456A, and N1457A (amino acid numbering beginning with the initiation Met codon). Mutations were confirmed by DNA sequencing.

COS-1 cells were transiently transfected with the indicated plasmid DNAs by diethylaminoethyl (DEAE) dextran using 2 μg/mL of DNA (*31*). At 60 h post-transfection, cells were rinsed with methionine-free Dulbecco's modified essential medium (DMEM) for 15 min and radiolabeled with [^{35}S]-methionine/cysteine (1000 Ci/mmol, Amersham Pharmacia Biotech, Inc.) in methionine-free DMEM for 25 min. Chase was performed for 0, 1, 4, or 12 h in medium containing a 10-fold excess of unlabeled methionine. At the indicated times, cells were harvested and lysed in IGEPAL CA-630 buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% IGEPAL CA-630 (Sigma, St. Louis, MO), 0.02% aprotinin, 1 mg/mL soybean trypsin inhibitor, and 1 mM phenylmethylsulfonylfluoride). Human and endogenous UGTs were immunoprecipitated from equal volumes of cell extracts with a rabbit polyclonal antibody made against native rat UGT (αN-350, kindly provided by Dr. Armando Parodi, Buenos Aires, Argentina). Antibody-protein immune complexes were adsorbed with protein A sepharose beads and washed sequentially with 1, 0.5, and 0.05% Triton X-100 in phosphate-buffered saline (PBS). Immunoprecipitates were treated with SDS-PAGE sample loading buffer for analysis by electrophoresis. Polyacrylamide gels were fixed with 30% methanol/10% acetic acid and treated with EN^{3}HANCE (Dupont-NEN, Boston, MA) for autoradiography. Western blot analysis was performed with a 1:200 dilution of a rabbit

polyclonal antibody made against native rat UGT (αGT333, kindly provided by Dr. Armando Parodi, Buenos Aires, Argentina) and developed by enhanced chemilumenescence (ECL, Amersham Pharmacia Biotech, Inc.).

Immunofluorescence Microscopy. COS-1 cells were transiently transfected by DEAE-dextran with 2 μg/mL HUGT1 or ΔHUGT1 (*31*). At 48 h post-transfection, cells were subcultured onto acid-washed coverslips and grown for an additional 16 h. Cells were fixed in a solution of 4% paraformaldehyde in PBS for 30 min, washed several times with 10 mM glycine in PBS, and permeabilized in a solution of 0.1% Triton X-100 in PBS-glycine for 5 min. Following several washes with PBS-glycine, cells were blocked with 100% goat serum for 45 min and exposed to primary antibody (either αGT333 at a 1:150 dilution, a mouse monoclonal αKDEL antibody (StressGen Biotechnologies Corp., Victoria, BC) at a 1:200 dilution, or both) in a solution of 2% goat serum in PBS-glycine for 45 min. Following several washes with PBS-glycine, cells were exposed to secondary antibody (either cy3-tagged goat anti-rabbit antibody (Jackson labs, Bar Harbor ME) at a 1:200 dilution, fluorescein isothiocyanate (FITC)-tagged goat anti-mouse antibody (Jackson labs) at a 1:200 dilution, or both) in a solution of 2% goat serum and 1:10 000 4′,6-diamidino-2-phenylindole (DAPI) in PBS-glycine for 45 min. Following a final series of washes with PBS-glycine, coverslips were mounted onto slides with Prolong solution (Molecular Probes, Eugene, OR). Mounting solution was allowed to dry overnight at 4 °C. Cells were viewed with an Olympus BX60 fluorescence microscope.

UGT Assay. COS-1 cells were cotransfected by DEAE-dextran with 1.5−2 μg/mL of the indicated plasmid expression vectors and 0.5−1 μg/mL of the puromycin resistance expression vector (either pEDpur (kindly provided by M. Davies, Genetics Institute Inc., Cambridge, MA) or pPUR (Clontech)). Where indicated, 10 μg/mL puromycin was added at 38−39 h post-transfection. After 30−36 h exposure to puromycin, cell extracts were collected in IGEPAL CA-630 lysis buffer or NP40 lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% IGEPAL CA-630 NP40 (Sigma), 0.02% aprotinin, 1 mg/mL soybean trypsin inhibitor, and 1 mM phenylmethylsulfonylfluoride) and assayed for protein concentration by the *DC* protein assay (BIO-RAD, Hercules, CA).

Samples were analyzed for glucosyltransferase activity as described (*9*, *23*), with slight modifications. Briefly, bovine thyroglobulin (Sigma Corp.), 20 mg/mL in 25 mM Tris-HCl, pH 8.0, was denatured for 4 h by dialysis at 23 °C against 8 M urea, 5 mM Tris-HCl, pH 7.5, and then warmed to 37 °C for 1 h. The urea was rapidly removed at 0 °C by dialysis against 25 mM Tris-HCl, pH 7.5, and kept at 0 °C until used in the assay. Standard assays contained in a 50 μL volume (exceptions noted in the text): 200 μg thyroglobulin, 2.5 μmol UDP-[^{14}C]-glucose (300 mCi/mmol, Dupont-NEN), 300 μM deoxynojirimycin, 5 mM $CaCl_2$, 25 mM Tris-HCl, pH 7.5, and cell lysate. Protein concentrations of cell lysates were kept constant within each experiment. When increasing volumes of cell extracts were used, lysis buffer was added to maintain the same detergent concentration. After incubation at 37 °C, samples were cooled to 0 °C and then 1 mL of 10% w/v trichloroacetic acid (TCA) was added. After 30 min, the precipitates were collected on GF/C glass microfiber filters (Whatman, Clifton), washed with 10% TCA, 95% ethanol, and acetone, the residue was dried, and the radioactivity measured. When RNase B (Sigma) served as substrate, it was denatured in 8 M urea at 60 °C for 2 h, reduced with 20 mM dithiothreitol, carboxymethylated with 100 mM iodoacetamide, and then dialyzed at 0 °C against 25 mM Tris-HCl, pH 7.5. To study the effect of divalent cations, incubation mixtures contained 1 mM EDTA (ethylenediaminetetraacetic acid) plus 10 mM $CaCl_2$, 10 mM $MgCl_2$, or 10 mM $MnCl_2$.

RESULTS

Isolation of cDNAs Encoding Human UDP-Glc:Glycoprotein Glucosyltransferase. A 438 base pair region of *Drosophila* UDP-Glc:glycoprotein glucosyltransferase, coding for residues (numbered 1296−1441) previously determined to have a high degree of homology to glycosyltransferases of other species (*23*), was used as a probe to screen a human fetal liver cDNA library. The overlapping clones obtained from several rounds of screening are depicted in Figure 1A. The combined HUGT1 sequence totals 4910 bp in length and encodes a 1555 residue protein with features supporting a soluble, ER-resident status (Figure 2). The amino-terminal residues underlined in Figure 2 comprise a region of high hydrophobicity that is predicted to function as the ER-targeting signal sequence with cleavage occurring after Ala42. The remainder of the protein sequence lacks a putative transmembrane domain, has a predicted molecular weight of 173 kDa, and has three potential N-glycosylation sites at residues 269, 536, and 1220. The locations of the N-linked glycosylation sites are not conserved in *Drosophila*, *C. elegans*, or *S. pombe* homologous sequences. There are thirteen cysteine residues present in the human sequence, five of which are conserved in the *Drosophila*, *C. elegans*, and *S. pombe* homologues. Four of these cysteine residues are in the proposed catalytic domain (described below) and likely form two disulfide loops. The carboxy-terminal four amino acids, REEL, represent a putative ER retrieval signal. The sequence REEL has been demonstrated to function in a manner analogous to the conserved mammalian retrieval signal, KDEL (*32*). The human sequence aligns well with homologous protein sequences from *Drosophila*, *C. elegans*, and *S. pombe*, both in terms of size and sequence conservation (Figure 1B). The region of highest identity among homologues resides within the final 300 amino acids; this portion is homologous to full-length glycosyltransferases from bacterial species, including RfaI and RfaJ proteins from *Escherichia coli* and *Salmonella typhimurium* and lgtC proteins from *Neisseria gonorrheae* and *Neisseria meningitidis* (*23*), and is, therefore, likely to contain the catalytic domain. A search of the GenBank EST database with HUGT1 sequence information led to the identification of a set of human EST sequences sharing homology with HUGT1. These ESTs were compiled and found to encode a partial protein sequence (termed HUGT2) with high identity to the carboxy-terminal portion of HUGT1 (residues 1102-end of HUGT1 and 1080-end of HUGT2, Figure 2). Overlapping clones hybridizing to HUGT2 were isolated from a human fetal liver cDNA library after several rounds of screening (Figure 1A).

The combined HUGT2 sequence totals 4828 bp in length and encodes a 1516 residue protein that, like HUGT1, has

features supporting a soluble, ER-resident status (Figure 2). The amino-terminal residues comprise a putative ER-targeting signal sequence with cleavage predicted to occur after Gly27. The remainder of the protein sequence lacks a putative transmembrane domain, has a predicted molecular weight of 172 kDa, and has four potential N-glycosylation sites at residues 256, 286, 920, and 950. The first of these glycosylation sites is conserved between the two human homologues, but none are conserved with the *Drosophila*, *C. elegans*, or *S. pombe* homologous sequences. There are eleven cysteine residues present in HUGT2, including the five conserved cysteines described for HUGT1. Four of the remaining HUGT2 cysteines are conserved with the HUGT1 sequence. The carboxy-terminal four amino acids, HDEL, represent a putative ER retrieval signal that is known to function in *Saccharomyces cerevisiae* (*33*). HUGT1 shares 55% identity with HUGT2 (Figure 1B). The alignment of HUGT2 with UGTs from *Drosophila*, *C. elegans*, and *S. pombe* found the highest identity within the terminal 300 amino acids, the proposed catalytic domain. HUGT2 shares overall identities with these species as follows: 40% with *Drosophila* UGT, 41% with the *C. elegans* homologue identified as Ce-1, 34% with the *C. elegans* homologue identified as Ce-2, and 30% with *S. pombe* UGT.

HUGT mRNAs Are Expressed In Many Tissues But Differ In Stress Inducibility. HUGT1 and HUGT2 mRNA expression were characterized by Northern blot hybridization analysis using poly $(A)^+$ mRNA isolated from multiple human tissues. The Northern blot was hybridized with sequences derived from the 3′-most regions of the two homologues. The identity shared by the two proteins falls off within the terminal 32 and 18 amino acids of HUGT1 and HUGT2, respectively, and a comparison of the corresponding nucleotide sequences demonstrated a high degree of sequence divergence that persists to the sites of polyadenylation (Figure 3A). Therefore, nucleotide fragments generated from the termini, encompassing partial coding sequence and most (HUGT2) or all (HUGT1) of the untranslated region (Figure 3A, underlined sequences), were used as probes. Three mRNA species, of approximate sizes 5.7, 8, and 10.6 kb, were detected in all tissues analyzed for expression of HUGT1 (Figure 3B). Additional mRNA species were detected, most notably the 3.1 and 6.5 kb signals in skeletal muscle, but likely represent cross-hybridization to the HUGT1 3′-untranslated region. This interpretation was supported by the absence of these signals from a hybridization to a 1 kb probe from the 5′-portion of HUGT1 (Figure 3B). Comparison to β-actin hybridization, a control for RNA loading, indicated that the highest level of expression was in the pancreas, while the lowest levels were in the lung and heart. The 3′ HUGT2 probe detected one mRNA species of approximately 5.1 kb (Figure 3B). Its pattern of tissue expression differed slightly from those of the HUGT1 mRNAs.

The proposed role of UGT in retaining unfolded proteins in the ER suggests that its expression might be induced under conditions that cause the accumulation of these substrates. The induction of UGT mRNAs in response to accumulation of unfolded protein in the ER was studied. COS-1 monkey cells were treated for increasing periods of time with either tunicamycin, an inhibitor of N-linked glycosylation, or the ionophore A23187, a reagent that causes efflux of Ca^{2+} from the ER and disrupts protein folding. Poly $(A)^+$ RNA was then isolated and analyzed by Northern blot hybridization. The 5′-HUGT1 probe detected COS-1 cell mRNAs corresponding to those found in multiple human tissues as well as an approximately 9.4 kb signal that was previously undetectable, possibly due to the lower resolution of the multiple tissue blot (compare Figure 4 and Figure 3B). Relative to β-actin, all UGT1 mRNA species in COS-1 cells were induced approximately 3−4-fold after 10 h treatment with either tunicamycin or A23187. In contrast, COS-1 cell UGT2 mRNA was not induced in response to tunicamycin and was only induced 2-fold in response to A23187 treatment (Figure 4). Under the same conditions, BiP mRNA, an extensively characterized ER-stress inducible mRNA (*34*, *35*), was induced 15- and 18-fold after a 10 h treatment with tunicamycin or A23187, respectively. Thus, the monkey homologue of HUGT1 is stress-inducible under these conditions, although to a much lesser extent than BiP, while the monkey homologue of HUGT2 is essentially noninducible.

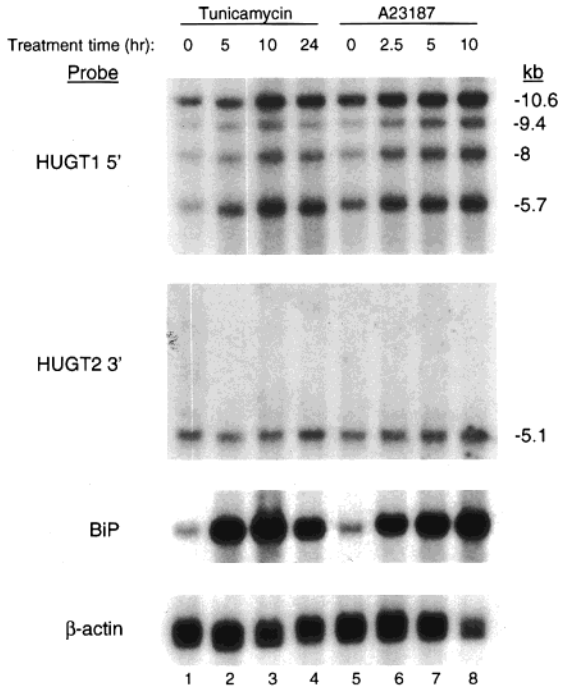


FIGURE 4: UGT1, and not UGT2, mRNA is induced by either tunicamycin or A23187 treatment. Poly $(A)^+$ RNA was extracted from total RNA collected from COS-1 cells treated with 10 μg/mL tunicamycin or 5 μM A23187 for the indicated times. A Northern blot containing this RNA was hybridized to the HUGT1 and HUGT2 probes described in Figure 3A. A BiP probe was used to demonstrate ER-stress induction, and a β-actin probe was used as a reference to quantify levels of induction.

HUGT1, And Not HUGT2, Encodes A Functional UGT Upon Expression In COS-1 Cells. cDNAs encoding full-length HUGT1 and HUGT2 were assembled and subcloned into a mammalian expression vector, pEDΔC. To measure the activity of HUGT1 and HUGT2, COS-1 cells were cotransfected with either expression vector in the presence of a vector encoding a *Streptomyces alboniger* protein, puromycin-*N*-acetyl transferase, that confers resistance to the antibiotic puromycin (Pur^R) (*36*). At 38 h post-transfection, cells were treated with puromycin for 30 h. Under these conditions, mock-transfected cells did not survive the puromycin selection, whereas approximately 20% of the cells

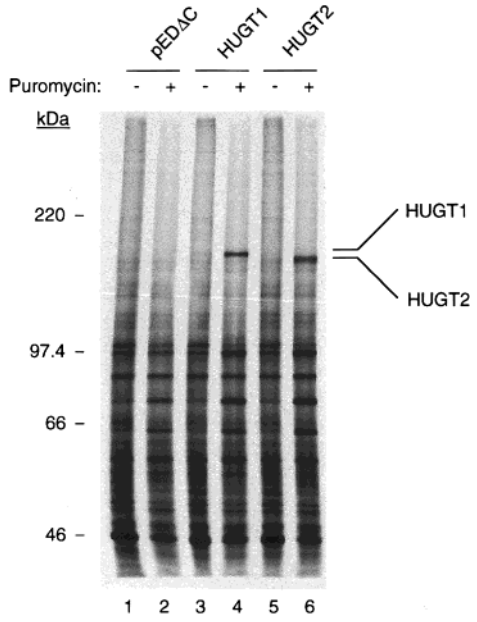


FIGURE 5: Synthesis of HUGT1 and HUGT2 in COS-1 cells. Equal numbers of total counts from extracts of pulse-labeled cells were analyzed directly by SDS-PAGE to measure synthesis of HUGT1 and HUGT2.

transfected with Pur^R survived, indicating significant selection for a relatively homogeneous population of transfected cells that received plasmid DNA. SDS-PAGE analysis of total cell extracts prepared from [^{35}S]-methionine pulse-labeled cells demonstrated that the level of HUGT synthesis was increased in both HUGT1- and HUGT2-transfected cells to a similar extent as determined by the presence of polypeptides migrating at approximately 175 kDa and 170 kDa, respectively (Figure 5). Upon selection with puromycin the signals for both HUGT1 and HUGT2 increased to a similar extent, indicating that both proteins were expressed at similar levels in the subpopulation of transfected cells. In addition, cell lysates analyzed for activity were subjected to Western blot analysis (data not shown). The steady-state levels of HUGT1 and HUGT2 mRNA were comparably increased upon puromycin selection to levels proportional to their rates of synthesis.

The assay for measuring the glucosyltransferase activity of the expressed HUGT constructs was optimized and used within the linear response range of each variable (data not shown). The incorporation of [^{14}C]Glc increased linearly with time up to 6 h and with increasing amounts of added cell extract containing 2−20 μg protein. Furthermore, increasing incorporation of glucose into thyroglobulin was observed when 45−350 μg of denatured thyroglobulin or 1.2−6.0 μmole UDP-[^{14}C]Glc were present. Denatured RNase B, used at 50−300 μg per reaction, was also labeled linearly by HUGT1. Parallel controls for the assay utilized purified *Drosophila* UGT (*23*).

Extracts from cells transfected with HUGT1 and selected with puromycin transferred [^{14}C]Glc from UDP-[^{14}C]Glc to denatured substrates RNase B and thyroglobulin approximately 6-fold and 8-fold greater, respectively, than control vector-transfected cells (Table 1A). In contrast to the findings with HUGT1, the activity of HUGT2 in transfected COS-1 cells did not increase compared to cells transfected with vector alone (Table 1A). Addition of cell extract from HUGT2-transfected cells to extract from HUGT1-transfected cells did not decrease the amount of activity detected for HUGT1 (Table 1B). These results indicate that not only was HUGT2 inactive but also it did not compete with HUGT1 for substrate binding to inhibit HUGT1 activity.

Table 1: Glucosyltransferase Activities of HUGT1 and HUGT2[a]

A. HUGT activity	fold endogenous activity: RNase B	fold endogenous activity: thyroglobulin
pEDΔC	1	1
HUGT1	6.4 ± 2.1	8.1 ± 3.4
HUGT2	1.2 ± 0.3	0.9 ± 0.1

B. competition assay	% HUGT1 activity
HUGT1	100
HUGT2	7.4 ± 0.3
HUGT1 + HUGT2	99.6 ± 3.0

C. HUGT1 substrate dependence	cpm
native thyroglobulin	48
denatured thyroglobulin	3680
HUGT1 divalent cation dependence	
1 mM EDTA	88
1 mM EDTA, 10 mM $CaCl_2$	2726
1 mM EDTA, 10 mM $MnCl_2$	513
1 mM EDTA, 10 mM $MgCl_2$	64

[a] All glucosyltransferase activities were measured in the presence of Pur^R selection. (A) Incorporation of [^{14}C]Glc into RNase B or thyroglobulin was determined for two independent transfection experiments per substrate. Fold endogenous activity was calculated relative to vector-transfected background. (B) The effect of HUGT2-transfected cell extract (4 μg total protein) on the glucosyltransferase activity present in HUGT1-transfected cell extract (4 μg total protein); RNase B was substrate for two independent transfections. The vector-transfected background counts were subtracted from the remaining values and the percentage activities were determined with respect to the activity of HUGT1 (arbitrarily set to 100%). (C) Glucosyltransferase activity in HUGT1-transfected cell extract was assayed as described in the Experimental Procedures, except where indicated, 200 μg denatured bovine thyroglobulin was replaced with 200 μg native bovine thyroglobulin, or 5 mM $CaCl_2$ was replaced with the components listed under divalent cation dependence. Although these results are representative of one experiment, duplication of the transfection and activity assay (again using thyroglobulin as substrate) gave qualitatively similar results.

Previous analyses of UGT demonstrated a requirement for denatured substrate and a cofactor preference among divalent cations (*9*, *21*, *23*). Correspondingly, the activity of overexpressed HUGT1 specifically required the thyroglobulin substrate to be denatured and displayed the expected ion preference for either Ca^{2+} or Mn^{2+}, and not Mg^{2+} (Table 1C). The results of Table 1 were verified with denatured RNase B in place of thyroglobulin (data not shown).

To determine if the HUGT1-dependent increase in UGT activity was actually due to the expressed HUGT1 protein, HUGT1 was mutagenized in order to destroy functional activity. Therefore, six consecutive amino acids, DQDLPN, from the highly conserved carboxy-terminal portion of the molecule were deleted in HUGT1 (residues 1452−1457, Figure 2). The second, third, and the sixth residues in this stretch are conserved among glucosyltransferases and homologous glycosyltransferases from human, *Drosophila*, *C. elegans*, *S. pombe*, *S. cerevisiae*, *Bacillus subtilis*, *E. coli*, and *S. typhimurium*. The first residue is conserved in all but the *S. cerevisiae* protein, a homologue for which glucosyltransferase activity has not been demonstrated.

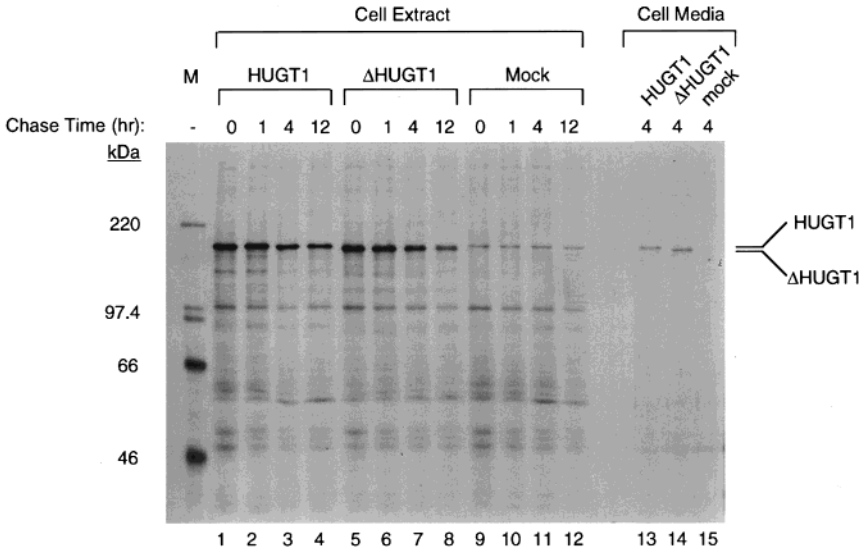


FIGURE 6: Pulse-chase analysis of wild-type and mutant HUGT1. COS-1 cells were mock-transfected or transfected with the HUGT1 or ΔHUGT1 expression vector, and at 60 h post-transfection, cells were labeled with [^{35}S]-methionine/cysteine for 25 min and chased in medium containing a 10-fold excess of unlabeled methionine for the indicated times. Overexpressed HUGT or endogenous COS-1 cell UGT proteins were immunoprecipitated from cell extracts or cell media. The cell extract portion of the autoradiograph represents a one day exposure to film, while the cell media portion represents a five day exposure. Mock represents cells that did not receive DNA.

COS-1 cells were transfected with either the wild-type (HUGT1) or mutant (ΔHUGT1) expression constructs and the synthesis and degradation of the proteins were measured by [^{35}S]-methionine/cysteine pulse-labeling and chase analysis. Immunoprecipitation of cell extracts with anti-rat UGT antibody and analysis by SDS-PAGE and autoradiography detected both HUGT1 and ΔHUGT1 polypeptides (Figure 6). Whereas HUGT1 comigrated with an endogenous COS-1 UGT at 175 kDa, ΔHUGT1 displayed a slightly faster mobility, consistent with the six amino acid deletion. The level of labeled endogenous UGT protein in mock-transfected cell extract remained relatively constant throughout the 12 h chase period (Figure 6; lanes 9−12). In contrast, the level of overexpressed HUGT1 or ΔHUGT1 decreased by approximately one-half by the 4 h chase time point (Figure 6; lanes 3,7), indicating that HUGT1 and ΔHUGT1 have similar half-lives. Immunoprecipitation of either HUGT1 or ΔHUGT1 from the 4 h chase-conditioned media detected small amounts of overexpressed protein in the media when the autoradiograms were over-exposed (Figure 6; lanes 13, 14). It is possible that the high level of expression exceeded the ER-retrieval capacity of the cell. Since the low amount of protein detected in the conditioned medium cannot account for the loss of protein from the cell extract, the decrease in the cellular level of overexpressed HUGT1 and ΔHUGT1 is likely due to intracellular degradation.

The functional activities of HUGT1 and ΔHUGT1 were measured by cotransfection of COS-1 cells in the presence of Pur^R. At 39 h post-transfection, cells were treated with puromycin for 36 h. Cell survival under these conditions was comparable to that seen above. SDS-PAGE analysis of extracts prepared from pulse-labeled cells demonstrated that HUGT synthesis was detected in both HUGT1- and ΔHUGT1-transfected cells and was prominent after puromycin selection (data not shown). Western blot analysis indicated that the steady-state level of ΔHUGT1 was lower than that of HUGT1 (Figure 7A; compare lanes 4 and 6). In addition, the steady-state level of HUGT1 was decreased by approximately one-half upon cotransfection with ΔHUGT1 (Figure 7A, compare lanes 4 and 8).

Cells transfected with HUGT1 (○) displayed a 9-fold greater rate of incorporation of [^{14}C]Glc from UDP-[^{14}C]-Glc into the denatured substrate thyroglobulin than cells transfected with the Pur^R vector alone (▲) (Figure 7B). The incorporation of [^{14}C]Glc was linear over the protein concentration range assayed and increased linearly with the time of incubation (Figure 7B and data not shown, respectively). Treatment of HUGT1-transfected cells with puromycin increased the rate of incorporation by approximately 3-fold (●) over unselected cells (○). Western blot analysis demonstrated that the level of increased activity in HUGT1-transfected cells approximately correlated with the increase in the steady-state amount of UGT over the endogenous level found in vector-transfected cells (data not shown), indicating a specific activity comparable to that of the endogenous COS-1 cell enzyme.

In contrast, COS-1 cells transfected with ΔHUGT1 displayed the same activity as cells transfected with the Pur^R vector alone either in the presence (×) or absence (■) of puromycin selection (Figure 7B), indicating that the deletion of the DQDLPN sequence from the carboxy-terminal region abolished the glucosyltransferase activity of the overexpressed protein. Significantly, the level of endogenous COS-1 cell UGT activity was not altered by overexpression of ΔHUGT1, indicating that the mutant protein did not act in a trans-dominant negative manner to inhibit the endogenous COS-1 cell UGT. Cotransfection of HUGT1 with ΔHUGT1 reduced the activity to approximately 50% (×) of that obtained by transfection of HUGT1 alone (●) (Figure 7C). This corresponded to the decrease in the HUGT1 steady-state level detected by Western Blot analysis (Figure 7A). Furthermore, ΔHUGT1 was detected migrating just under the HUGT1 polypeptide (Figure 7A; lane 8), supporting the interpretation that ΔHUGT1 was not behaving in a trans-dominant negative manner. Repetition of the cotransfection (Table 2) confirmed the reproducibility of the above findings.

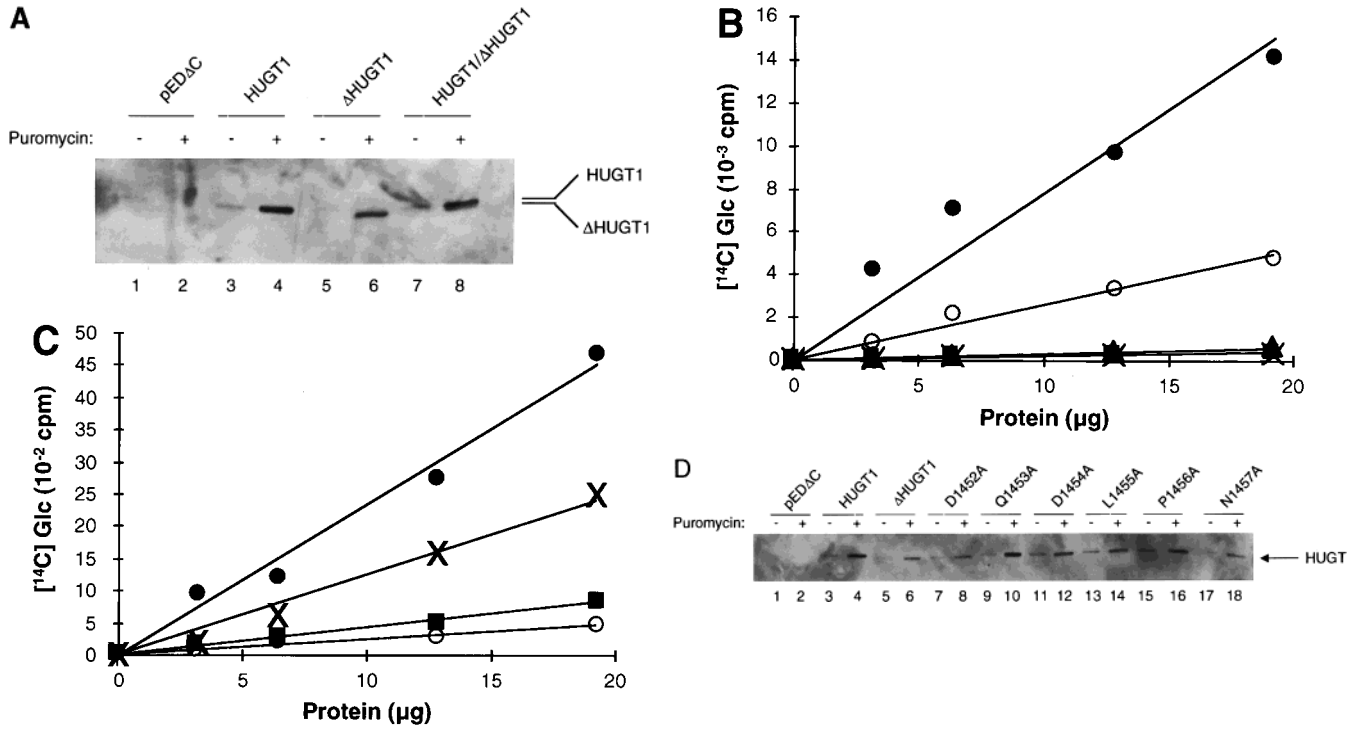


FIGURE 7: Expression and activity of site-directed HUGT1 mutants. Panel A: Western blot analysis of lysates from transfected cells (50 μg of protein per sample) using an αUGT primary antibody. Panel B: Cell extract protein concentration curve for incorporation of [^{14}C]Glc from UDP-[^{14}C]Glc into denatured bovine thyroglobulin for a Pur^R cotransfection: (●) HUGT1/Pur^R selected; (○) HUGT1/Pur^R unselected; (×) ΔHUGT1/Pur^R selected; (■) ΔHUGT1/Pur^R unselected; (□) Mock Pur^R selected; (▲) Mock and Pur^R unselected. Panel C: Cell protein extract concentration curve for incorporation of [^{14}C]Glc into denatured bovine thyroglobulin for a HUGT1 and ΔHUGT1 cotransfection: (●) HUGT1, (○) ΔHUGT1, (×) HUGT1/ΔHUGT1, and (■) mock. Although the numbers presented here represent single experiments, they were corroborated by additional independent experiments (Table 2A). Panel D: Western blot analysis of wild-type and mutant HUGT1-transfected cell lysates using αUGT primary antibody. Equal amounts of protein (50 μg) from cell extract were loaded into each lane.

Table 2: Mutational Analysis of HUGT1 Catalytic Domain[a]

construct(s)	% HUGT1 activity
HUGT1	100
ΔHUGT1	0 ± 0
HUGT1/Δ*H*UGT1	48.3 ± 3.8
D1452A	0.3 ± 0.4
Q1453A	3.7 ± 2.6
D1454A	0.1 ± 0.2
L1455A	1.6 ± 1.2
P1456A	40.6 ± 5.2
N1457A	6.8 ± 0.7

[a] HUGT1-, ΔHUGT1-, and HUGT1 point mutant-transfected COS-1 cells were selected with puromycin, cell extracts were prepared, and UGT activity was determined as described in Materials and Methods. The vector-transfected background counts were subtracted from the remaining values and the percentage activities were determined relative to the activity of wild-type HUGT1 (again set to 100%). Data represent analysis of two independent transfection experiments utilizing denatured thyroglobulin as substrate.

Addition of increasing amounts of extract from ΔHUGT1-transfected cells to extract from HUGT1-transfected cells did not decrease the amount of activity detected for HUGT1 (data not shown). This result indicates that ΔHUGT1, similar to HUGT2, cannot compete for substrate binding under conditions of limiting substrate.

Amino Acids Critical For HUGT1 Function. To identify amino acid residues required for HUGT1 function, point mutations to alanine were made for each of the six amino acids. All vectors were cotransfected with Pur^R. SDS-PAGE analysis of extracts prepared from pulse-labeled cells demonstrated that the level of HUGT synthesis was increased to a similar extent in all cells transfected with the mutant HUGT constructs (data not shown). Cell extracts from puromycin-selected cells were analyzed for activity (Table 2). Although the point mutants D1452A, D1454A, and L1455A were detectable by western blot analysis (Figure 7D; lanes 8, 12, and 14, respectively), their expression did not increase glucosyltransferase activity above background (Table 2). While the Q1453A mutant displayed negligible glucosyltransferase activity (Table 2), its steady-state level was equal to or even greater than that of the wild-type protein (Figure 7D; lane 10), supporting the conclusion that it was inactive as well. In contrast, point mutants P1456A and N1457A displayed reduced activity in comparison to wild-type (Table 2) and were expressed at lower steady-state levels than HUGT1 (Figure 7D; compare lanes 16 and 18 with lane 4). Even after correction for the differences in UGT levels between samples, P1456A and N1457A were significantly less active than the wild-type protein.

Expressed HUGT1 Localizes To The Endoplasmic Reticulum. Immunofluorescence microscopy was used to identify the intracellular location of the HUGT1 overexpressed in COS-1 cells. Transiently transfected cells were stained with either a polyclonal rabbit antibody against rat UGT, a mouse monoclonal antibody against a KDEL containing peptide, or both, and then treated with a secondary anti-rabbit cy3-tagged antibody, a secondary anti-mouse FITC-tagged antibody, or both, respectively (Figure 8). While the subpopulation of HUGT1-transfected cells stained with a much

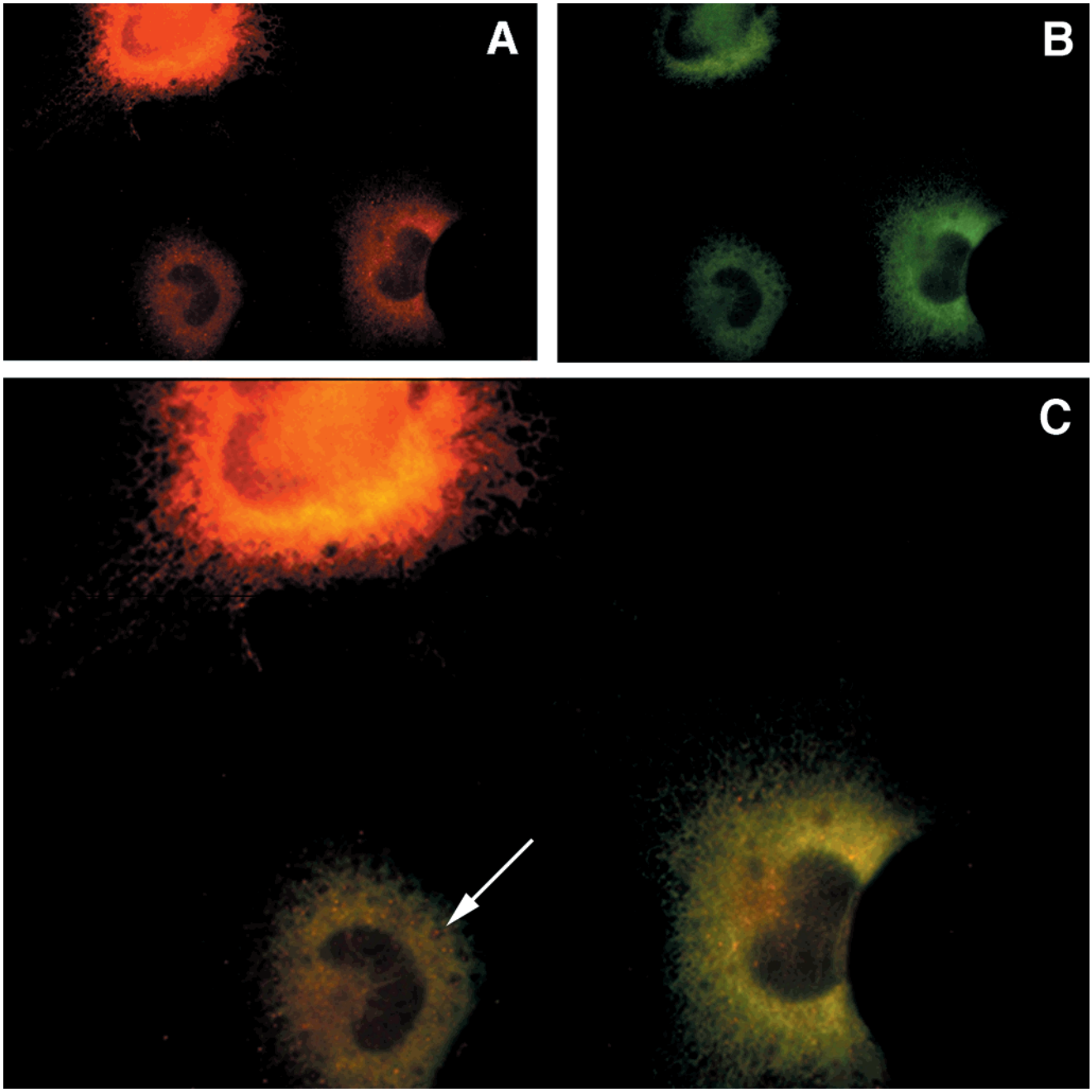


FIGURE 8: Localization of overexpressed HUGT1 in transfected COS-1 cells. COS-1 cells were transfected with the HUGT1 expression vector, subcultured onto coverslips at 48 h post-transfection, and analyzed by immunofluorescence 16 h later as described in Experimental Procedures. Panel A: 60X magnification of cells treated with αUGT primary antibody and cy3-tagged secondary antibody. Panel B: 60X magnification of cells treated with αKDEL primary antibody and FITC-tagged secondary antibody. Panel C: Merged view of panels A and B, to identify colocalization. Arrow designates a representative cy3-fluorescence that does not colocalize.

greater intensity than nearby untransfected cells (Figure 8A), the pattern of staining in both cell types was similar, indicating that the overexpressed UGT protein was not mislocalized. A comparison of Panel A with Panel B demonstrates that this staining pattern was similar to that resulting from exposure of the same cells to the αKDEL antibody, which recognizes BiP and two additional proteins believed to be GRP94 and Hsp47 (*37*). A merging of the separate projection images obtained from cy3-only and FITC-only fluorescence produced a yellowed coloration of the costained cells, indicating colocalization of UGT with the ER-resident chaperones (Figure 8C). Thus, although HUGT1 was very highly overexpressed, the majority of the overexpressed protein did colocalize with the ER-resident KDEL proteins. Perinuclear localization was evident upon comparison with DAPI nuclear staining of the same cell population (data not shown). Although endogenous UGT colocalized with the ER marker proteins, there were also visible red spots (Figure 8C; example indicated by arrow). Similar structures were recently described for endogenous UGT in Chinese hamster ovary cells and were proposed to represent ER exit sites where intermediate compartment vesicles are formed (*38*). Analysis of ΔHUGT1-transfected COS-1 cells indicated a pattern of localization similar to that displayed by HUGT1 (data not shown). The localization of HUGT1 and ΔHUGT1 to the ER is consistent with results demonstrating that the overexpressed and endogenous proteins are sensitive to endoglycosidase H (data not shown),

an enzyme that cleaves high mannose oligosaccharides prior to transport through the Golgi compartment.

DISCUSSION

We isolated one cDNA, HUGT1, encoding a functional human UDP-Glc:glycoprotein glucosyltransferase and established the transcription of a closely related homologue, HUGT2. The conclusion that HUGT1 encodes a functional UGT is supported by numerous criteria. Its deduced amino acid sequence predicts a signal peptide and an ER retrieval signal and shares extensive sequence homology with *Drosophila*, *C. elegans*, and *S. pombe* homologues. Transient expression of the cDNA in COS-1 cells produced a glycosylated protein that was localized to the ER and reacted with an anti-rat UGT antibody. HUGT1-transfected cells displayed an increased UDP-Glc:glycoprotein glucosyltransferase activity, including a specificity for denatured, but not native, glycoprotein substrates, and a cofactor requirement for Ca^{2+} or Mn^{2+} (*9*, *21*, *23*). Four single amino acid mutations in a region of highest identity, the proposed catalytic domain, extinguished the glucosyltransferase activity.

The fundamental characterization of UGT by Parodi and co-workers demonstrated that this enzyme, expressed in most eukaryotic species, has the unique ability to selectively glucosylate unfolded glycoproteins (*39*, *9*, *21*). UGT genes from *Drosophila* and *S. pombe* have been described and putative *C. elegans* homologues exist in the Genbank database (*23*, *40*). HUGT1 is the first mammalian homologue for which full-length cDNA sequence information has been elucidated. We draw several conclusions from the comparisons that are now possible. The human protein shares identity over its entire length with identified UGTs from other species, with the highest identity lying within the C-terminal 300 residues. Several bacterial glycosyltransferases share homology over their entire lengths to the C-terminal region of the UGTs. Therefore, this region likely contains the catalytic domain. The less well-conserved regions of the protein might provide substrate recognition. This type of conservation is exemplified by the hsp70 family members, which share highly conserved catalytic domains and display extensive divergence within their peptide-binding domains (*1*).

Our results are the first to demonstrate that recombinant-derived HUGT1 has the properties of UGTs characterized from *Drosophila*, rat, and *S. pombe*. Immunoprecipitation of HUGT1 from transfected COS-1 cell lysates with anti-rat UGT antibody yielded a single 175 kDa species that comigrated with a COS-1 cell protein, the presumed endogenous COS-1 cell UGT. This result supports the existence of a COS-1 cell protein homologous to HUGT1. Furthermore, expression of HUGT1 protein in COS-1 cells increased the UGT activity in transfected cells in proportion to the amount of HUGT1 protein produced. Deletion of the C-terminal conserved hexapeptide DQDLPN in ΔHUGT1 destroyed the increased activity, indicating that the elevated activity resulted from expression of HUGT1 and was not due to a transfection-stimulated, endogenous COS-1 cell activity. Results from overexpression of ΔHUGT1 suggest that although the ΔHUGT1 protein is enzymatically inactive, it cannot act as a trans-dominant negative inhibitor of either endogenous COS-1 cell or transfected HUGT-1 transferase activity. Because gel filtration of rat liver UGT under native conditions indicated that its size is 270 kDa, it was suggested that the protein forms dimers (*9*). If homodimerization is an element of HUGT1 function, then either ΔHUGT1 has no effect on HUGT1 activity upon dimerization, or the six amino acid deletion prohibits dimerization of ΔHUGT1 with HUGT1. Furthermore, ΔHUGT1 did not compete with HUGT1 for substrate binding, as addition of increasing amounts of extract from ΔHUGT1-transfected cells failed to decrease the glucosyltransferase activity of separately expressed HUGT1.

To further substantiate the above findings and to rule out a general disruption of HUGT1-folding by the six amino acid deletion in ΔHUGT1, alanine point mutants at each of the six residues were studied. Mutation of any one of four adjacent residues, Asp1452, Gln1453, Asp1454, or Leu1455, destroyed functional activity. Remarkably, mutation of the highly conserved sixth residue, Asn1457, did not fully abolish catalysis. Recently, the *Drosophila* and *S. pombe* homologues of HUGT1 were placed in galactosyltransferase Family B by homology to bacterial proteins involved in lipopolysaccharide core biosynthesis (*41*). A hydrophobic cluster alignment of Family B members established the presence of a conserved motif corresponding to the hexapeptide described above, DQDXXN (*41*). An additional motif, identified as DXD and present in all families of galactosyltransferases but one (*41*), is also conserved in Family B and, therefore, present in HUGT1 and its homologues. It was proposed that coordination of the UDP-sugar requires a divalent cation and that the DXD motif of the transferase provides the binding site for this metal ion (*42*). Adopting the proposal that the DXD motif represents the site of UDP-Glc binding, we suggest that the DQDLPN region of HUGT1 is involved in recognition of the substrate to which the glucose is transferred. Previous studies demonstrated that the *N*-acetylglucosamine residue linked to the consensus asparagine is essential for UGT substrate recognition (*21*). In addition to this primary key interaction, exposed hydrophobic patches on the incorrectly folded glycoprotein might then stimulate the catalytic activity of HUGT1. We propose that ΔHUGT1 and the corresponding nonfunctional point mutants are defective in *N*-acetylglucosamine binding and that this interaction mediates the primary, essential contact with the substrate. This would explain the inability of ΔHUGT1 to compete with wild-type HUGT1. Further analysis of the site-directed mutants of HUGT1 should provide an opportunity to test this hypothesis.

Partial sequence information for a second potential UDP-Glc:glycoprotein glucosyltransferase, HUGT2, was obtained from the compilation of a set of human ESTs too dissimilar to be representative of HUGT1. These ESTs encode a partial open reading frame with high identity to the presumed catalytic domain of HUGT1 and contain a distinct putative ER retrieval signal, HDEL. The full-length sequence was isolated and shown to share 55% identity with HUGT1. HUGT1 and HUGT2 were expressed at comparable levels upon transfection of COS-1 cells. Though they additionally shared similar electrophoretic mobilities by SDS-PAGE analysis, HUGT2 could not be immunoprecipitated with the anti-rat antibodies that recognized HUGT1, and more importantly, failed to show UGT activity under the conditions

optimized for HUGT1. Northern blot analysis showed that HUGT2 mRNA is expressed in multiple human tissues, although in a distinctively different pattern than HUGT1 mRNA. These results support that the HUGT2 mRNA sequence is not derived from the HUGT1 gene nor is it an artifact from assembling of EST clones. Furthermore, while ER stress stimulated the expression of the monkey homologue of HUGT1 mRNA, ER stress did not increase the expression of the monkey homologue of HUGT2 mRNA. Several explanations are possible for the lack of HUGT2 glucosyltransferase activity. First, HUGT2 may have a substrate specificity different from that of HUGT1, making it inactive in the assays employed here. For example, a UDP-glucose glucosyltransferase was recently characterized that transfers glucose to fucose residues present on glycoproteins (*43*). The region most divergent between HUGT1 and HUGT2 lies within the amino termini of the proteins and may represent a substrate interaction site that differs between the two homologues. In support of this idea is the intriguing fact that monkey UGT2 mRNA was not stress-inducible, a possible reflection of the inability of HUGT2 to recognize unfolded glycoproteins as substrates. Second, it is possible that HUGT2 has diverged from HUGT1 so that the UGT activity has been lost. HUGT2 has four amino acid residues within the proposed catalytic domain that differ from the residues conserved in all other UGTs (Leu1231, Arg1306, Arg1384, and Tyr1436, as identified by an asterisk in Figure 2). The divergence of these four residues was verified by analysis of the EST database, demonstrating that these changes were not due to a cloning or sequencing error. Third, it is possible that the conserved HUGT2 catalytic domain might be involved in the transfer of a sugar other than glucose. For example, several of the bacterial transferases with homology to the catalytic region of the UGTs are responsible for the transfer of UDP-galactose rather than UDP-glucose (*41*). Fourth, it is possible that HUGT2 shares with HUGT1 the conserved regions required for recognition of a malfolded N-linked glycoprotein, but then initiates a course of action differing from that of HUGT1. For example, in place of reglucosylation, it might direct the glycoprotein toward proteolytic degradation. This would be important if the mannose acceptor sites for glucose were lost before a native fold was achieved (*22*). Finally, an intriguing possibility is that HUGT2 might require an associated protein or chaperone for catalytic activity. For example, ERp57 is a protein disulfide isomerase homologue that displays increased activity in the presence of either calnexin or calreticulin (*44*). We are presently testing the validity of these hypotheses.

In summary, in humans, there exist two related genes with different expression characteristics of which only one encodes the expected UGT activity. Notably, in addition to the presence of two UGT homologues in *C. elegans*, our examination of the current murine EST database led to the identification of two sets of homologous overlapping ESTs. One set is highly identical both in sequence and length to HUGT1, while the other is more similar to HUGT2. Thus, the presence of two related, UGT-like genes appears to be conserved in higher eukaryotic species, although only one homologue from any species has been identified as an active UGT. These studies provide an essential foundation for uncovering human pathologies in this fundamental ER detector of abnormal glycoprotein folding, and the discovery of HUGT2 points to the presence of additional genomic information dedicated to sensing glycoprotein folding.

ACKNOWLEDGMENT

We thank Dr. Armando Parodi for generously providing anti-rat UGT antibodies. We also thank Dallas Kroon for technical assistance and Joe Nowak for his Photoshop expertise in generation of the immunofluorescence figures. Finally, we thank Dr. Steven Pipe for his advice in the editing of the manuscript.

REFERENCES

1. Gething, M. J., and Sambrook, J. (1992) *Nature 355*, 33.
2. Lodish, H. F., Hong, N., Snider, M., and Strous, G. J. (1983) *Nature 304*, 80.
3. Hurtley, S. M., and Helenius, A. (1989) *Annu. Rev. Cell Biol. 5*, 277.
4. Kornfeld, R., and Kornfeld, S. (1985) *Annu. Rev. Biochem. 54*, 631.
5. Hettkamp, H., Legler, G., and Bause, E. (1984) *Eur. J. Biochem. 142*, 85.
6. Brada, D., and Dubach, U. C. (1984) *Eur. J. Biochem. 141*, 149.
7. Trombetta, E. S., Simons, J. F., and Helenius, A. (1996) *J. Biol. Chem. 271*, 27509.
8. Ganan, S., Cazzulo, J. J., and Parodi, A. J. (1991) *Biochemistry 30*, 3098.
9. Trombetta, S. E., and Parodi, A. J. (1992) *J. Biol. Chem. 267*, 9236.
10. Ou, W.-J., Cameron, P. H., Thomas, D. Y., and Bergeron, J. J. M. (1993) *Nature 364*, 771.
11. Le, A., Steiner, J. L., Ferrell, G. A., Shaker, J. C., and Sifers, R. N. (1994) *J. Biol. Chem. 269(10)*, 7514.
12. Pind, S., Riordan, J. R., and Williams, D. B. (1994) *J. Biol. Chem. 269*, 12784.
13. Loo, T. W., and Clarke, D. M. (1994) *J. Biol. Chem. 269*, 28683.
14. Pipe, S. W., Morris, J. A., Shah, J., and Kaufman, R. J. (1998) *J. Biol. Chem. 273*, 8537.
15. Labriola, C., Cazzulo, J. J., and Parodi, A. J. (1995) *J. Cell. Biol. 130(4)*, 771.
16. Bergeron, J. J., Brenner, M. B., Thomas, D. Y., and Williams, D. B. (1994) *Trends. Biochem. Sci. 19*, 124.
17. Michalak, M., Milner, R. E., Burns, K., and Opas, M. (1992) *Biochem. J. 285*, 681.
18. Vassilakos, A., Michalak, M., Lehrman, M. A., and Williams, D. B. (1998) *Biochemistry 37*, 3480.
19. Trombetta, E. S., and Helenius, A. (1998) *Curr. Opin. Struct. Biol 8*, 587.
20. Hebert, D. N., Foellmer, B., and Helenius, A. (1995) *Cell 81*, 425.
21. Sousa, M., and Parodi, A. J. (1995) *EMBO J. 14*, 4196.
22. Liu, Y., Choudhury, P., Cabral, C. M., and Sifers, R. N. (1999) *J. Biol. Chem. 274*, 5861.
23. Parker, C. G., Fessler, L. I., Nelson, R. E., and Fessler, J. H. (1995) *EMBO J. 14*, 1294.
24. Sousa, M. C., and Parodi, A. J. (1996) *Cell Mol. Biol. Paris 42*, 609.
25. Hammond, C., Braakman, I., and Helenius, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91(3)*, 913.
26. Kozak, M. (1987) *Nucleic Acids Res. 15*, 8125.
27. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York.
28. Ting, J., Wooden, S. K., Kelleher, K. S., Kritz, R., Kaufman, R. J., and Lee, A. S. (1987) *Gene 55*, 147.
29. Cleveland, D. W., Lopata, M. A., MacDonald, R. J., Cowan, N. J., Rutter, W. J., and Kirschner, M. W. (1980) *Cell 20*, 95.

30. Kaufman, R. J., Davies, M. V., Wasley, L. C., and Michnick, D. (1991) *Nucleic Acids Res. 19*, 4485.
31. Kaufman, R. J. (1989) in *Methods in Enzymology: Gene Expression Technology* (Goeddel, D., Ed.) pp 487−511.
32. Haugejorden, S. M., Srinivasan, M., and Green, M. (1991) *J. Biol. Chem. 266(10)*, 6015.
33. Semenza, J. C., Hardwick, K. G., Dean, N., and Pelham, H. R. (1990) *Cell 61*, 1349.
34. Kozutsumi, Y., Segal, M., Normington, K., Gething, M. J., and Sambrook, J. (1988) *Nature 332*, 462.
35. Lee, A. S. (1987) *Trends Biochem. Sci. 12*, 20.
36. Lacalle, R. A., Ruiz, D., and Jimenez, A. (1991) *Gene 109*, 55.
37. Cribb, A. E., Pohl, L. R., Spielberg, S. P., and Leeder, J. S. (1997) *J Pharmacol. Exp. Ther. 282*, 1064.
38. Cannon, K. S., and Helenius, A. (1999) *J. Biol. Chem. 274*, 7537.
39. Trombetta, S. E., Bosch, M., and Parodi, A. J. (1989) *Biochemistry 28*, 8108.
40. Fernandez, F. S., Trombetta, S. E., Hellman, U., and Parodi, A. J. (1994) *J. Biol. Chem. 269*, 30701.
41. Breton, C., Bettler, E., Joziasse, D. H., Geremia, R. A., and Imberty, A. (1998) *J. Biochem. (Tokyo) 123*, 1000.
42. Parodi, A. J. (1999) *Biochim. Biophys. Acta 1426*, 287.
43. Moloney, D. J., and Haltiwanger, R. S. (1999) *Glycobiology 9*, 679.
44. Zapun, A., Darby, N. J., Tessier, D. C., Michalak, M., Bergeron, J. J., and Thomas, D. Y. (1998) *J. Biol. Chem. 273*, 6009.
45. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng. 10*, 1.

BI9916473