Primary structure of the cytosolic β-glucosidase of guinea pig liver

William S. HAYS*, Steven A. JENISON†‡, Takashi YAMADA†§, Andrzej PASTUSZYN* and Robert H. GLEW*‖
Departments of *Biochemistry and †Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM 87131, U.S.A.

The cytosolic β-glucosidase (EC 3.2.1.21) present in the livers of mammalian species is distinguished by its broad specificity for sugars and its preference for hydrophobic aglycones. We purified the cytosolic β-glucosidase from guinea pig liver and sequenced 142 amino acid residues contained within 12 trypsin digest fragments. Using degenerate oligonucleotide primers deduced from the peptide sequences, a 622 bp cytosolic β-glucosidase cDNA was amplified by reverse-transcriptase PCR, using total guinea pig liver RNA as template. The 'rapid amplification of cDNA ends (RACE)' method [Frohman (1993) Methods Enzymol. **218**, 340–356] was used to synthesize the remaining segments of the full-length cDNA. The complete cDNA contained 1671 nucleotides with an open reading frame coding for 469 amino acid residues. The amino acid sequence deduced from the cDNA sequence included the amino acid sequences of all 12 trypsin digest fragments derived from the purified enzyme. Amino acid sequence analysis indicates that the guinea pig liver cytosolic β-glucosidase is a Family 1 β-glycosidase and that it is most closely related to mammalian lactase-phlorizin hydrolase. These results suggest that the cytosolic β-glucosidase and lactase-phlorizin hydrolase diverged from a common evolutionary precursor.

INTRODUCTION

In mammals, three prominent β-glucosidases have been well characterized: lysosomal glucocerebrosidase, intestinal lactase-phlorizin hydrolase (LPH), and the cytosolic β-glucosidase found predominantly in the liver and kidneys. The metabolic roles of both LPH and glucocerebrosidase are known, and deficiencies in these two enzymes are associated with specific diseases. LPH hydrolyses dietary lactose (EC 3.2.1.108) and glucosylceramide (EC 3.2.1.62); LPH deficiency causes lactose intolerance [1–3]. Glucocerebrosidase (EC 3.2.1.62) hydrolyses glucosylceramide derived from endogenous membrane glycolipids; deficiency of this enzyme results in Gaucher disease [4,5]. Neither a metabolic role nor an associated disease state have been described for the cytosolic β-glucosidase [6].

Several kinetic and physical features distinguish the cytosolic β-glucosidase from LPH and glucocerebrosidase. Unlike LPH and glucocerebrosidase, the cytosolic β-glucosidase is freely soluble and is not glycosylated [7–9]. While it accepts a variety of sugar glycosides as non-physiological substrates, the cytosolic β-glucosidase does not hydrolyse the natural substrates of either LPH or glucocerebrosidase [7,9]. Furthermore, it will not hydrolyse a number of β-glycosides known to be produced endogenously in mammals [8,9].

The inability to identify an endogenous substrate for the cytosolic β-glucosidase has led to the hypothesis that the enzyme may serve to cleave dietary xenobiotic glycosides [10]. Supporting this idea is the observation that the enzyme efficiently hydrolyses a variety of plant glycosides, including amygdalin and prunasin [10,11]. However, the cytosolic β-glucosidase has not yet been shown to be required for the metabolic disposal of xenobiotic glycosides. Thus there is insufficient evidence to either exclude the presence of an endogenous substrate or to confirm a role for the cytosolic β-glucosidase in xenobiotic transformation.

In this report, we describe the partial amino acid sequence and the complete cDNA sequence of the cytosolic β-glucosidase from guinea pig liver. Analysis of these sequences indicates that the enzyme is a Family 1 β-glucosidase having the greatest homology with LPH. We discuss the implications of these data for the structure and function of this enzyme.

MATERIALS AND METHODS

Materials and reagents

Livers from Dunkon–Hartley guinea pigs were supplied by Rockland, Inc. (Rockland, PA, U.S.A.). Substrates for glucosidase assays, histone H1 and dipalmitoyl-phosphatidylserine were obtained from Sigma (St. Louis, MO, U.S.A.). α-[^{35}S]dATP and γ-[^{32}P]ATP were purchased from Amersham (Arlington Heights, IL, U.S.A.). The Sequenase 2.2 DNA sequencing kit from Amersham-USB was used for the determination of cDNA nucleotide sequences. Recombinant human casein kinase 2 (CK2), rat brain protein kinase C (PKC), and 1-stearoyl-2-arachidonoyl-*sn*-glycerol were purchased from Calbiochem (La Jolla, CA, U.S.A.). All other chemicals were of reagent grade.

Enzyme purification

All purification steps were performed in a standard buffer containing 10 mM sodium phosphate, pH 6.0, and 1 mM β-mercaptoethanol (β-ME), with specific modifications as noted

Abbreviations used: DTT, dithiothreitol; IEF, isoelectric focusing; LPH, lactase-phlorizin hydrolase; PEG, poly(ethylene glycol); RACE, rapid amplification of cDNA ends; β-ME, β-mercaptoethanol; RT-PCR, reverse-transcriptase PCR; PKC, protein kinase C; CK2, casein kinase 2; TFA, trifluoroacetic acid.

‡ Present address: NM Department of Health, 525 Camino de los Marquez, Suite 1, Santa Fe, NM 87501, U.S.A.
§ Present address: Department of Obstetrics and Gynaecology, Hokkaido University School of Medicine, N15 W7 Kita-Ku Sapporo, 060 Japan.
‖ To whom correspondence should be addressed.

below. Unless otherwise stated, all procedures were performed at 4 °C or on ice. Ten guinea pig livers (183 g) were homogenized in buffer supplemented with protease inhibitors [1 mM EDTA, 1 mM β-ME, 0.02 % NaN_3, 1 μg/ml leupeptin, 6 μg/ml pepstatin, 5 μg/ml aprotinin, 17 mM PMSF, 1 % (v/v) isopropanol] using a Waring blender and brought to a final volume of 380 ml. The homogenate was centrifuged at 100000 ***g*** at 4 °C for 75 min. The supernatant was precipitated with 20 % (v/v) poly(ethylene glycol) (PEG) 4000 (final concentration), and centrifuged at 2500 ***g*** for 30 min. β-Glucosidase activity was removed from the PEG supernatant by batch elution from a DEAE-cellulose column, using buffered 1 M NaCl as the eluent.

The eluent was dialysed against buffer to a final conductivity of 2500 μS and then passed through a 0.45-μm-pore-size filter. The filtrate was loaded on to a Waters Protein-Pak DEAE-5PW steel column (21.5 mm × 150 mm) using a Waters 650 Advanced Protein Purification System (Waters, Milford, MA, U.S.A.). The column was eluted with a linear gradient of 0 to 0.15 M NaCl; β-glucosidase activity was eluted in a single peak at 0.06 M NaCl. The pooled fractions from the DEAE column were brought to a final concentration of 1.0 M $(NH_4)_2SO_4$ in the standard buffer, and passed through a 0.45 μm syringe filter. The filtrate was loaded on to a phenyl-substituted hydrophobic interaction column (Waters/Shodex PH-814; 8 mm × 75 mm) and eluted with a decreasing linear gradient from 1.0 to 0 M $(NH_4)_2SO_4$. The peak activity fractions were concentrated and desalted using a Centricon 30 ultrafiltration device (Amicon, Beverly, MA, U.S.A.), and injected on to a Waters DEAE-5PW anion-exchange column (7.5 mm × 75 mm) equilibrated in the standard buffer at pH 7.0. The column was eluted with a linear gradient of 0 to 0.2 M NaCl, with β-glucosidase activity being eluted at 0.08 M NaCl. The enzyme was stored in 50 % (v/v) ethylene glycol at −20 °C, where it is stable for at least 1 year.

Enzyme assays

Assays for β-glucosidase activity were performed as described previously [12] using fluorescence assays. One unit of enzyme activity equals one μmol of substrate converted/min.

Protein determinations

Protein concentrations were measured using the dye-binding method of Bradford [13] with BSA as the standard.

Peptide mapping and amino acid sequencing

Purified β-glucosidase (100 μg) in 50 % (v/v) ethylene glycol was concentrated with a Centricon 30 ultrafiltration device, washed with 2.0 ml of 50 mM Tris, pH 8.1, and concentrated to a final volume of 220 μl. The enzyme was heated to 99 °C for 10 min, cooled to 2 °C, and incubated at 37 °C for 48 h with 2 μl of a 10 % (w/v) solution of trypsin (Sigma). The trypsin digest was stored at −20 °C for later analysis. Prior to separation by HPLC, the digest was thawed and β-ME was added to a final concentration of 0.1 M. The solution was vortexed briefly, then incubated at 37 °C for 15 min. The trypsin digest was injected on to a 5 μm, 4.9 mm × 25 mm C_{18} reversed-phase HPLC column (Alltech, Deerfield, MI, U.S.A.) equilibrated with Solvent A [5 % CH_3CN, 0.1 % trifluoroacetic acid (TFA)] at a flow rate of 1.0 ml/min. The column was eluted with Solvent B (95 % CH_3CN, 0.1 % TFA) according to the following linear gradient programme: 0–5 min, 100 % A; 5–127 min, 0–55 % B; 127–132 min, 55–100 % B; 132–142 min, 100 % B. The absorbance of the eluent was monitored at 214 nm, and fractions were collected manually. Fractions selected for sequencing were lyophilized in a Speed Vac (Savant, Farmingdale, NY, U.S.A.) and resuspended in 50 μl of water. An aliquot (20 μl) of the concentrated fraction was injected into a Beckman LF3000 automated microsequencer (Beckman, Fullerton, CA, U.S.A.).

Amino acid analysis

Amino acid analysis was performed on a sample (9 μg) of purified enzyme using a Waters Pico-Tag system according to established methods [14].

SDS/PAGE and isoelectric focusing (IEF)

Aliquots of purified enzyme (0.43 μg) in 50 % ethylene glycol were lyophilized, brought to 8 μl in 10 mM Tris/HCl/1 mM EDTA/5 % (w/v) SDS/0.01 % (w/v) Bromophenol Blue/1.5 M β-ME, and heated to 100 °C for 5 min. Electrophoresis was performed with a Pharmacia Phastgel system using 12.5 % homogenous media (Pharmacia, Piscataway, NJ, U.S.A.). For IEF, an aliquot (0.34 μg) of enzyme was focused using the Phastgel system and IEF 4–6.5 media. The IEF and SDS/PAGE gels were stained with $AgNO_3$ using the automated protocols provided by the manufacturer.

cDNA cloning and sequencing

A sample (0.5 g) of liver from a freshly killed adult guinea pig was homogenized in 5 ml of Tri-Reagent (MRC, Cincinnati, OH, U.S.A.), and total RNA was extracted according to the method of Chomczynski and Sacchi [15]. Total liver RNA (5 μg) was used as the template for the synthesis of first-strand cDNA, using Superscript reverse transcriptase (Gibco-BRL, Bethesda, MD, U.S.A.). Priming was performed with oligo-$dT_{12–17}$ primers, and reaction conditions were exactly as recommended by the manufacturer. PCR was performed using 1 μl aliquots of first-strand cDNA as template with hemi-nested primers corresponding to the coding sequences of tryptic peptides 89 and 126 (see Figure 3) [16]. The outer amplification included 5 pmol of primers 89S and A126C, 20 mM Tris/HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 2.5 units of *Taq* polymerase (Gibco-BRL), in a final volume of 50 μl. Amplification was performed for 45 cycles according to the following programme: 94 °C for 1.5 min; 45 °C for 2 min; 72 °C for 2 min. A final extension cycle was performed at 72 °C for 5 min. The reaction mixture from the outer amplification was diluted 1:20 with water, and an aliquot (1 μl) was used as the template for hemi-nested PCR with primers 89S and A126N (see Figure 3). Amplification was conducted under the same conditions and thermal cycler parameters as the outer amplification, with a final reaction volume of 100 μl. The single 622 bp amplification product was gel-purified and ligated into pCRII (Invitrogen, San Diego, CA, U.S.A.) as described previously [17]. A single clone (A1) was isolated, and the nucleotide sequence was determined by double-stranded dideoxy sequencing methods [18].

3′- and 5′-directed RACE (rapid amplification of cDNA ends)

The 'RACE' technique was used to amplify cDNAs from the 3′ and 5′ ends of the mRNA [19]. 3′-Directed RACE was performed exactly as described by Frohman [19] using the primer Q_T for first-strand cDNA synthesis. The sequences of all primers used for reverse-transcription PCR (RT-PCR) and RACE are shown in Figure 3; the primer pairs are shown with their orientations relative to the cDNA in Figure 4 (upper panel). The outer amplification employed the primers LS and Q_O (100 nM of each) according to the following thermal cycler programme: 94 °C for

Table 1 Summary of the purification of cytosolic β-glucosidase

Units of enzyme activity are expressed in μmol of substrate hydrolysed per min.

Step	Volume (ml)	Protein (mg)	Activity (μmol/min)	Specific activity (units/mg)	Fold purification	Yield (%)
Crude homogenate	380	6150	42.0	0.00683	(1)	(100)
100000 ***g*** supernatant	465	4700	47.8	0.0102	1.49	(100)
20% PEG supernatant	765	464	31.8	0.0686	10.0	75.7
DE52 (batch)	265	134	30.5	0.228	33.4	72.6
DEAE-5PW, pH 6.0	245	13.5	19.7	1.46	214	46.9
Hydrophobic affinity	47	1.25	6.65	5.32	779	15.8
DEAE-5PW, pH 7.0	7	0.293	4.7	16.0	2340	11.2

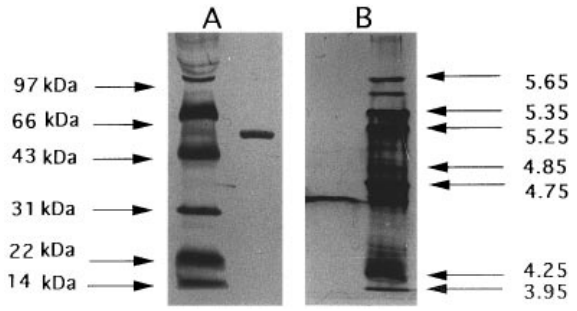
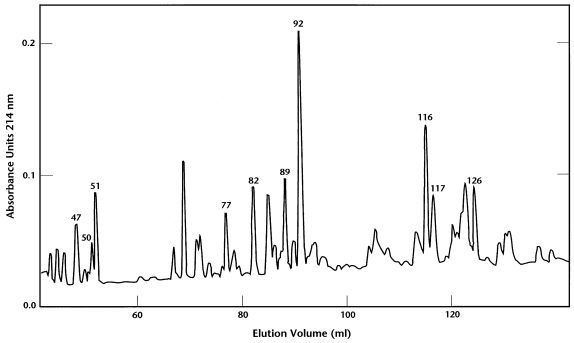


Figure 1 SDS/PAGE and IEF of the purified cytosolic β-glucosidase

The purity of the enzyme was assessed by SDS/PAGE (**A**) and IEF (**B**) gels as described in the Materials and methods section. Staining was with $AgNO_3$. The sizes of molecular mass markers in kDa (Bio-Rad low-range molecular mass markers; Hercules, CA, U.S.A.) are indicated to the left of (**A**); the apparent molecular mass of the enzyme is 53 kDa. The positions of the isoelectric point standards (Pharmacia LKB #1860-202 pI markers; Piscataway, NJ, U.S.A.) are indicated in pH units to the right of (**B**); the apparent pI of the enzyme is 4.6.

1.5 min; 54 °C for 1 min; and 72 °C for 40 min; followed by 30 cycles of the same programme with the extension step at 72 °C reduced to 2 min. A final extension step at 72 °C was performed for 5 min. The product of the outer amplification was diluted 1:20 with water, and an aliquot (1 μl) used as template for nested PCR. Nested PCR used the primer pair MS and Q_I (100 nM of each) for 30 cycles of the following programme: 94 °C for 1.5 min; 54 °C for 1 min; and 72 °C for 2 min; followed by an extension step at 72 °C for 5 min. The resulting PCR products were gel-purified and ligated into pCRII. Recombinant transformants were selected as described [17]. Clone B1 represented the complete 3′ end of the cDNA, including the polyadenylation signal.

5′-Directed RACE was performed using first-strand cDNA that was primed with the primer RAS from the 3′ end of clone A1 (see Figure 4, upper panel). Two first-strand reactions were combined (40 μl total) and desalted using a Centricon 30 and 1 × TE buffer (10 mM Tris/HCl, pH 7.8, 1 mM EDTA). The cDNA was resuspended in a final volume of 34 μl of 0.1 × TE. Aliquots (11 μl) were resuspended in a final volume of 20 μl containing 100 mM potassium cacodylate, pH 7.2, 2 mM $CoCl_2$, 0.2 mM dithiothreitol (DTT) and 15 units of terminal deoxynucleotidyl transferase (Gibco-BRL). The mixture was incubated at 37 °C for 5 min, 65 °C for 5 min, and then placed on ice. PCR was performed with 1 μl of the terminal transferase reaction as template. A combination of the primers Q_T (3.5 nM) and Q_O (100 nM) was used in the sense orientation and the primer MAS



#47	V	P	Y	T	S	A	K							
#51	G	V	D	Y	Y	N	K							
#77	W	E	X	F	R									
#82	F	S	I	S	W	S	R							
#89	N	Q	T	G	D	V	A	C	G	S	Y	T	L	W
#116	E	A/S	E/D	L	G	I	L	Q	D	A	E	I	E	L	F
#117	G	M/D	V/P	S	L/N	S/S	L/V	F	C	I	W	P	Q	P	E
#126	A	I/D	N/G	F/D	Q/Y	F/P	D	F	F	A	K	P	I	F	I
#50A	S	Q	I	A	S										
#92A	A	W	H	S	Y	D	S	L	F						
#92B	G	T	A	D	F	F	A	V	Q	Y	Y	T			
#92C	G	P	C	V	W	D	T	F	T	H	Q	G	G	E	

Figure 2 Peptide mapping and amino acid sequencing

Upper panel: chromatogram of the separation of trypsin-digested cytosolic β-glucosidase on a C18 reversed-phase HPLC column. The numbered peaks represent fractions yielding the sequences shown in the panel below. Lower panel: amino acid sequences of the tryptic peptides isolated by HPLC above. Peptides 47, 51, 77, 82, 89, 116, 117 and 126 were single sequences obtained from the respective fractions identified in the chromatogram shown in the upper panel. Peptides 50A, 92A, 92B and 92C were present as a mixture of peptides in their respective fractions. The sequences of these four peptides were deduced later from the cDNA sequence (see Figure 5). X = indeterminate amino acid assignment; the third residue in peptide 77 was determined to be a cysteine by inspection of the cDNA sequence.

(100 nM) used in the antisense orientation, with the final reaction volume being 50 μl. The thermal cycler programme was as follows: first cycle: 94 °C for 1.5 min, 45 °C for 2 min, and 72 °C for 40 min; second cycle: 94 °C for 1.5 min, 54 °C for 2 min, and 72 °C for 40 min; cycles 3–33: 94 °C for 1.5 min, 45 °C for 2 min, and 72 °C for 2 min. A final extension step was conducted at 72 °C for 5 min. Nested PCR (100 μl final) was conducted using the primers Q_I and LAS (100 nM each) and 1 μl of the outer amplification reaction (diluted 1:20 with water) as template. Thermal cycler parameters were as follows: 94 °C for 1.5 min, 54 °C for 1 min, and 72 °C for 2 min, 35 cycles total. PCR reaction conditions were otherwise identical to those described above. A single broad band of PCR product (approx. 400 bp) was produced, ligated into pCRII, and cloned as described above. The clones generated (C1–C32) were heterogeneous in length, differing in the extent of the 5′ cDNA sequence they contained. The longest clone (C19) contained the presumed translation initiation codon and 56 nucleotides of 5′ untranslated region.

Direct sequencing of PCR products

The nucleotide sequences of both strands of the cDNA were determined by direct sequencing of PCR products amplified from first-strand cDNA. A series of overlapping PCR products were generated that spanned the region from base 55 to base 1621 in the sense strand, and from base 57 to base 1553 in the antisense strand (see Figure 4, lower panel). PCR products were labelled with biotin at the 5′ end of one of the primers and purified for sequencing using streptavidin-coated magnetic beads (Dynal, Lake Success, NY, U.S.A.) [20]. Each segment of each strand was sequenced at least twice, using as template the products of separate PCR amplifications. The sequence of the 5′ untranslated region from base 44 to 54 was determined by sequencing three separate clones isolated during the 5′-RACE procedure. The sequence from base 1 to 43 was determined from the sequence of the longest 5′-RACE clone (C19).

In vitro phosphorylation assays

Phosphorylation of the cytosolic β-glucosidase by PKC was attempted *in vitro* using purified rat brain PKC. A concentrated buffer stock was prepared [0.4 M Tris, pH 7.3, 0.4 M NaCl, 2 mM $CaCl_2$, 0.2 mg/ml phosphatidylserine, 40 μg/ml diacylglycerol], sonicated in a water bath for 10 min, and placed on ice. Purified β-glucosidase (0.4 μg) was mixed in a final volume of 20 μl with 5 μl of concentrated buffer, 20 μM ATP, 10 mM $MgCl_2$, 0.9 μCi of [γ-^{32}P]ATP, and 90 m-units of PKC. Control reactions included PKC alone, PKC plus histone H1 (4 μg), and cytosolic β-glucosidase alone. The reactions were incubated for 30 min in a 37 °C water bath, then dialysed for 2 h against 1 litre of water as described [21]. The dialysate was lyophilized in a Speed Vac, and prepared for either SDS/PAGE or IEF analysis as described above, with Coomassie Blue staining. Autoradiograms were exposed for 2 h and compared with the stained gels.

Phosphorylation of the purified β-glucosidase *in vitro* by CK2 was investigated by combining the β-glucosidase (0.4 μg) in a final volume of 20 μl with 20 mM Mes, pH 6.9, 130 mM KCl, 10 mM $MgCl_2$, 4.8 mM DTT, 25 μM ATP, 1 μCi of [γ-^{32}P]ATP, and 0.2 m-unit of CK2. Control reactions included CK2 alone, CK2 plus casein (4 μg), and β-glucosidase alone. Reactions were incubated at 37 °C for 15 min in a water bath, followed by precipitation with 2 μl of 0.015% (w/v) sodium deoxycholate and 2 μl of 100% (w/v) trichloroacetic acid. Precipitates were spun briefly in a microfuge and the pellet washed with 20 μl of water. The pellet was resuspended in 5 μl of gel loading buffer and analysed by SDS/PAGE as described above. Autoradiograms were exposed for 48 h and compared with the Coomassie Blue-stained gels.

```
A126N     GCI AAI AAI TCI AAI TGI AAI TTI AT
A126C     TCI TAI CCI TCI ATI AAI ATI GGI TT
89S       ACI GGI GAI GT(GCT) GC(ACT) TG(CT) GG
Q_T       CCA GTG AGC AGA GTG ACG AGG ACT CGA GCT
              CAA GCT TTT TTT TTT TTT TTT T
Q_O       CCA GTG AGC AGA GTG ACG
Q_I       GAG GAC TCG AGC TCA AGC
LS        TCC TGG TCA CGT CTG TTA C
LAS       GTA ACA GAC GTG ACC AGG A
MS        GAA GAC CAA GGT GGA TGG T
MAS       ACC ATC CAC CTT GGT CTT C
RAS       AAG TGA TAG GGA CAC CAT AC
```

Figure 3 Sequences of primers used for RT-PCR and RACE

Primers A126N and A126C correspond to the antisense complements of the N- and C-termini of peptide 126. Primer 89S corresponds to the sense orientation of peptide 89. Primers Q_T, Q_O, and Q_I were identical to those described by Frohman for RACE [19].

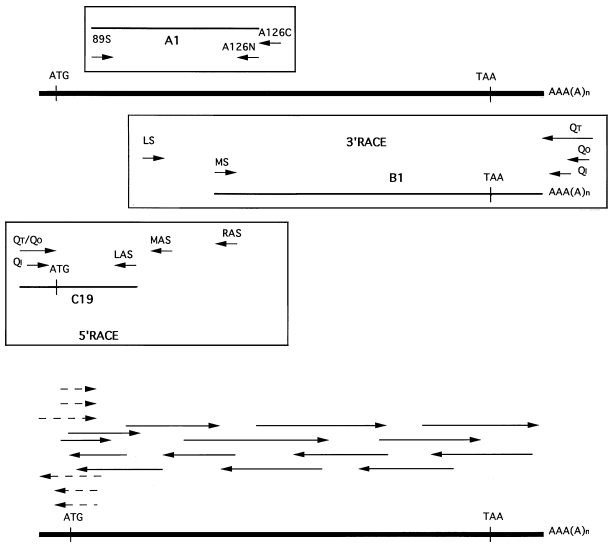


Figure 4 Cloning and sequencing strategies

Upper panel: the generation of clone A1 using the degenerate primers shown in Figure 3 is indicated in the upper box. The generation of the RACE clones B1 and C19 is indicated in the lower boxes. The sizes and locations of clones and primers are shown to scale with reference to the full-length mRNA, shown by the heavy black line. Vertical staggering of the primer pairs indicates nested PCR amplifications: upper and outer sets of primers represent outer amplifications, while lower and inner primer pairs represent nested amplifications. Primers Q_T and RAS were used to prime the first-strand reactions for clones B1 and C19 respectively. Lower panel: sequencing of the cDNA. Leftward arrows represent antisense-strand sequencing; rightward arrows represent sense-strand sequencing. Solid arrows indicate direct sequencing of PCR product; broken arrows represent sequence obtained from cloned PCR products.

Sequence analysis

Amino acid sequence analysis and alignments were performed using the Wisconsin Package (GCG, Madison, WI, U.S.A.). Database searches were performed using the NCBI BLAST algorithm [22]. Sequences identified by BLAST with scores of 150 or greater were selected for further analysis. Individual sequences were analysed for known structural motifs using MOTIFS, pairwise sequence alignments performed using GAP, and multiple sequence alignment performed using PILEUP.

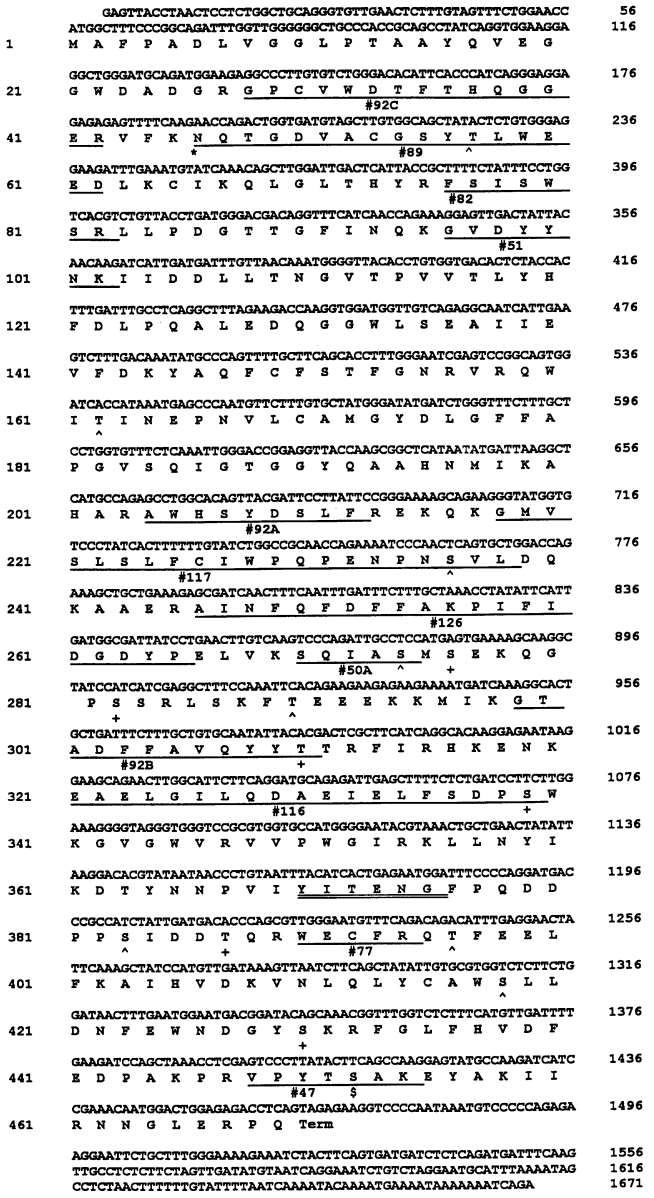


Figure 5 cDNA and amino acid sequences of the cytosolic β-glucosidase

The amino acid sequence numbers are specified along the left-hand side of the sequence; nucleotide sequence numbers are shown to the right. Underlined amino acid sequences represent the sequences of the tryptic peptides shown in Figure 2(lower panel). (*) = consensus N-glycosylation site; (^) = consensus phosphorylation site for CK2; (+) = consensus site for phosphorylation by PKC; (×) = consensus phosphorylation sequence for both CK2 and PKC. The doubly underlined sequence at amino acid residues 370–375 represents the consensus active site sequence (see the text).

RESULTS

Protein purification and peptide sequencing

The protocol employed to purify the cytosolic β-glucosidase is summarized in Table 1. The specific activity of the pure enzyme was 16.0 μmol/mg of protein per min, which is 2-fold greater than the highest value previously reported for the enzyme from guinea pig liver [12]. The purity of the enzyme was assessed on silver-stained SDS/PAGE and IEF gels. The apparent molecular mass was 53 ± 1 kDa, and the isoelectric point was 4.6 (Figure 1). Attempts to sequence the intact enzyme were unsuccessful, indicating a blocked N-terminus. Purified enzyme (100 μg) was subjected to exhaustive trypsin proteolysis, and the resulting peptides were separated by reversed-phase HPLC (Figure 2, upper panel). Eight of the resultant fractions were pure and yielded single sequences; an additional four fractions yielded a mixture of sequences (Figure 2, lower panel). Comparisons of the peptide sequences with the GENBANK, Brookhaven, and SWISS-PROT databases indicated that peptides 51, 82, 89 and 126 were homologous with LPH from human and rabbit intestine [23], while peptides 47, 77, 116 and 117 were unique.

cDNA cloning and sequencing

The elucidation of the full primary structure of the cytosolic β-glucosidase was achieved in two phases. First, overlapping fragments of the full-length cDNA were cloned and sequenced using RT-PCR. In the second phase, the nucleotide sequences of the cloned fragments were confirmed by direct sequencing of RT-PCR products amplified from the first-strand cDNA. This redundancy was necessary because sequence obtained from a single PCR-generated clone is more likely to reflect random errors introduced by *Taq* polymerase than is sequence information obtained directly from PCR products [24].

Table 2 Amino acid composition of cytosolic β-glucosidase

The amino acid compositions predicted from the cDNA sequence and measured with the purified enzyme are shown. The amino acid composition of the purified enzyme was determined as described in the Materials and methods section. Values for the acid/amide pairs Asp/Asn and Glu/Gln are reported as the sums D + N and E + Q for the composition determined from the pure enzyme. The accuracy of this method is ±10%. Abbreviation: ND, not determined.

Residue	Predicted	Observed
A	30	35
C	8	ND
D	32	**
N	21	**
D + N	53	55
E	29	**
Q	24	**
E + Q	53	58
F	32	34
G	36	39
H	9	9
I	28	25
K	28	31
L	34	36
M	6	ND
P	24	25
R	20	21
S	24	24
T	23	24
V	26	26
W	14	ND
Y	21	21

We isolated the initial clone by performing RT-PCR with degenerate oligonucleotide primers derived from the sequences of tryptic peptides 89 and 126 (Figures 2 and 3). The sense (89S) and antisense (A126C/A126N) orientations of the RT-PCR primers were designed on the basis of the relative positions of their homologous sequences in LPH. Peptide 89, from which primer 89S was developed, is homologous with amino acids 946–956 in domain III and 1419–1429 in domain IV of human prepro-LPH. The corresponding regions for peptide 126 were residues 1146–1165 and 1620–1639 of domains III and IV respectively. {The sequence of pro-LPH contains three fully repeated domains of approx. 450 amino acids. The two C-terminal domains (III and IV) are retained in the final mature protein as functionally independent β-glucosidase activities [23].} First-strand cDNA, generated from total liver RNA with an oligo-$dT_{12–17}$ primer, was used as template. A single 622 bp amplimer was cloned and sequenced. The nucleotide sequence of the clone, designated A1, contained a single full-length open reading frame that included the coding sequences of peptides 89, 82, 51, 117 and 126.

Using oligonucleotide primers derived from the sequence of clone A1, the 5′ and 3′ ends of the cDNA were amplified and cloned by the RACE procedure described by Frohman [19]. 3′-directed RACE was accomplished using the primer pairs Q_O/LS (outer) and Q_I/MAS (nested), with first-strand cDNA primed with Q_T as the template for the outer amplification. Five PCR

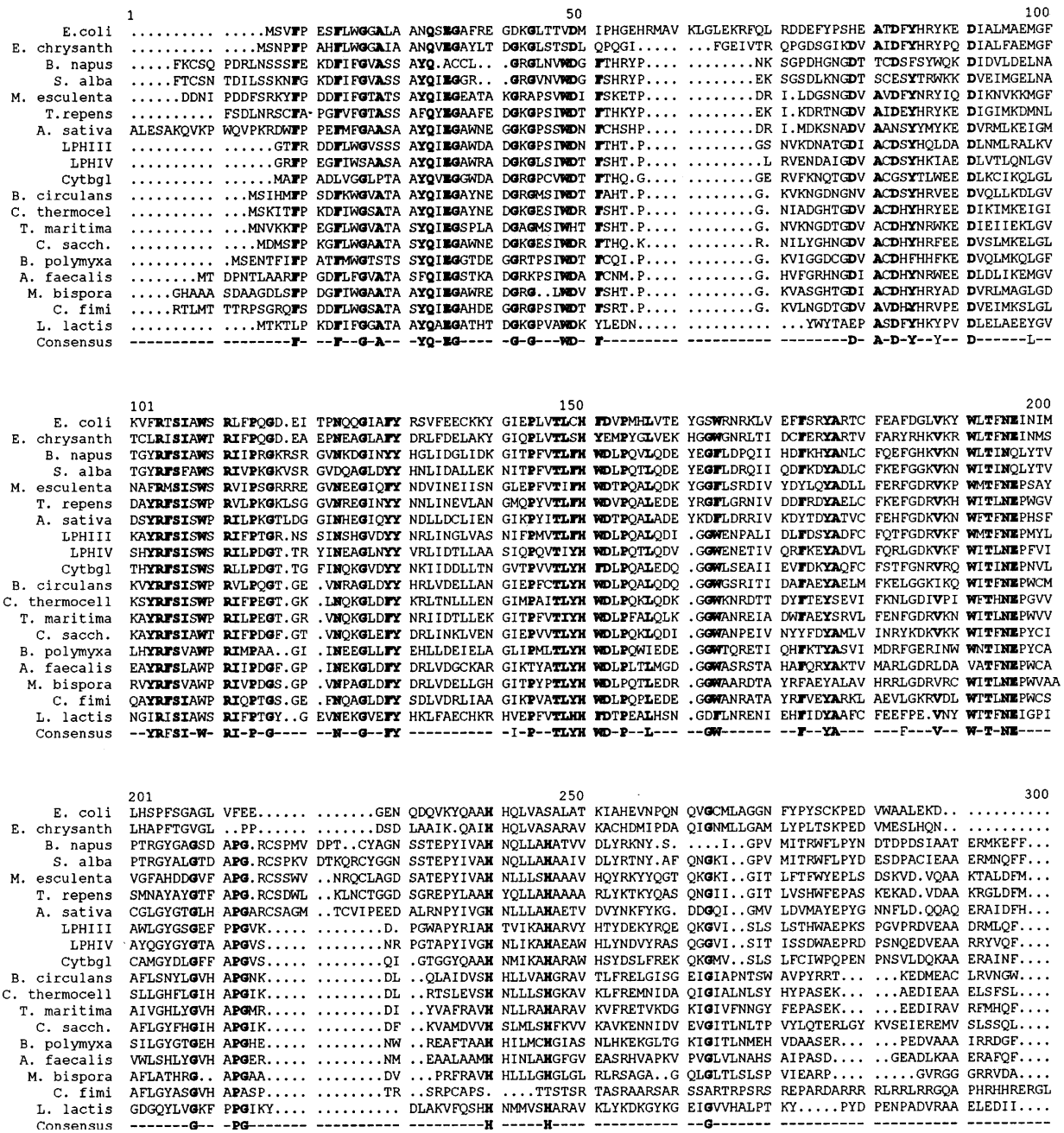


Figure 6 For legend see facing page

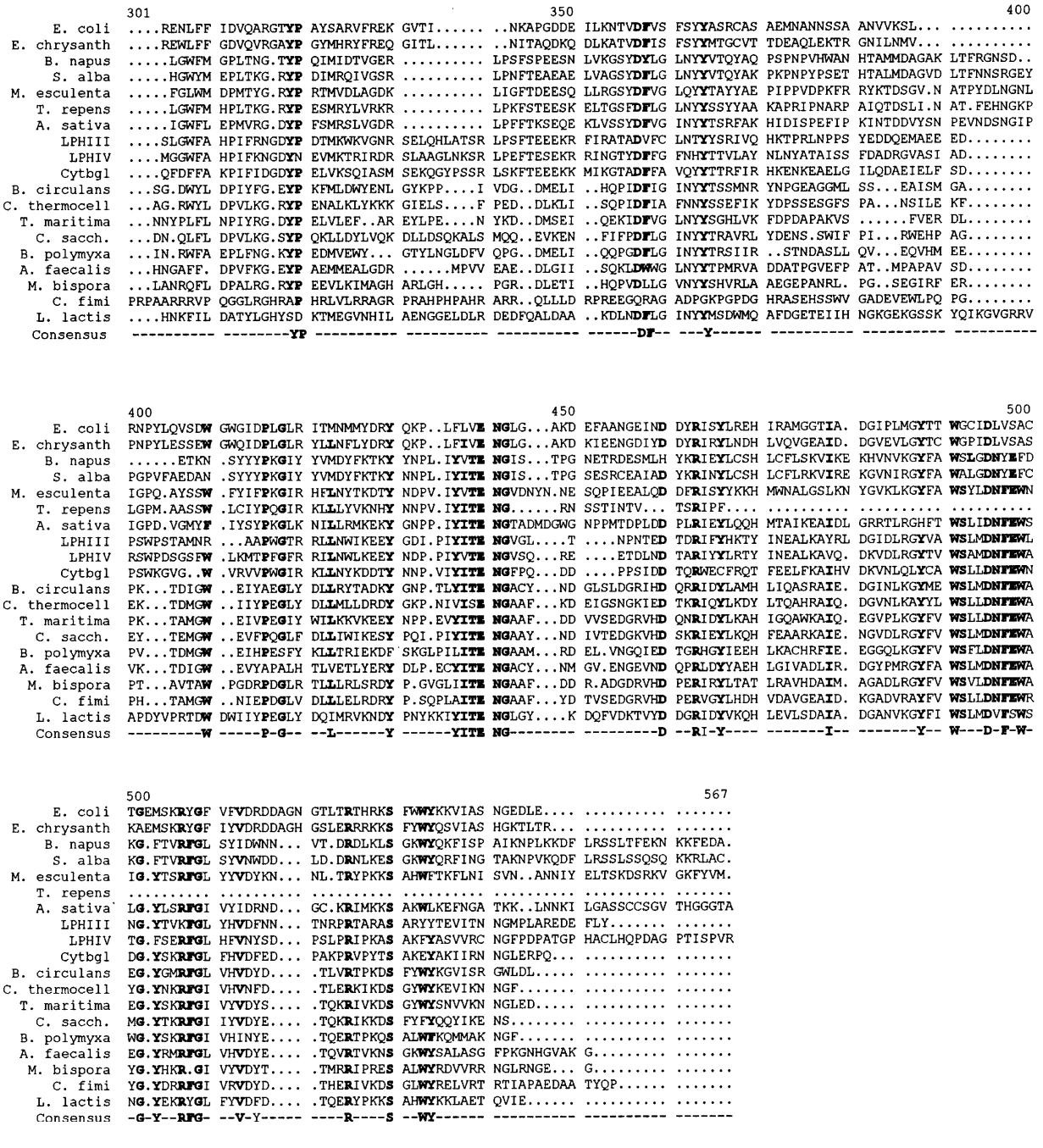


Figure 6 Alignment of the amino acid sequences of the cytosolic β-glucosidase and selected β-glucosidases

The sequences of the β-glucosidases from *Brasica napus* (B. napus)(S19146), *Sinapis alba* (S. alba)(S19149)[EC 3.2.3.1], *Manihot esculenta* (M. esculenta)(S23940), *Trifolium repens* (T. repens)(S16580), *Avena sativa* (A. sativa)(X78433), human LPH domains III (lactase, [EC 3.2.1.108]) and IV (glucosylceramidase, [EC 3.2.1.62]) (LPHIII and LPHIV)(S01168), *Microspora bispora* (M. bispora)(A48949), and *Cellumonas fimi* (C. fimi)(M98649) were truncated as appropriate at their N- and C-termini to exclude transmembrane domains and other non-homologous regions from the alignment. A consensus sequence was calculated at selected positions with well-conserved sequences. Amino acids and consensus sequences identical among 16 or more sequences are shown in bold-face type. Gaps in the alignment are indicated by dots (.). The identities of the other β-glucosidases are as follows: Cytbgl = cytosolic β-glucosidase from guinea pig liver; E. coli = *Escherichia coli* (S27553)[EC 3.2.1.86]; E. chrysanthemi = *Erwinia chrysanthemi* (C42603); B. circulans = *Bacillus circulans* (A48969); C. thermocellum = *Clostridium thermocellum* (S17215); T. maritima = *Thermotoga maritima* (S41561); C. saccharolyticum = *Caldocellum saccharolyticum* (S03813); B. polymyxa = *Bacillus polymyxa* (JW0038); A. fecaelis = *Agrobacterium faecalis* (A28673); L. lactis = *Lactococcus lactis* (M28357)[EC 3.2.1.85]. [GenBank or SwissProt accession numbers for each sequence are indicated in parentheses; the EC numbers for enzymes other than β-glucosidases (e.g. EC 3.2.1.21) are indicated in brackets where appropriate.]

products, ranging from 200 to 1100 bp, were produced and cloned. The largest clone (B1) included the stop codon and the polyadenylation signal, and overlapped at its 5′ end with the sequence of clone A1. 5′-Directed RACE was performed in a two-step procedure. The first-strand cDNA was generated using the primer RAS derived from the nucleotide sequence of clone A1. A poly(A) tail was added to the first-strand cDNA with terminal deoxynucleotidyl transferase. Fully nested PCR was performed,

using as the outer primers a mixture of Q_T and Q_O in the sense orientation and the primer MAS in the antisense orientation. The primer pair Q_I/LAS was used for the inner amplification. A broad band of PCR products of approx. 400 bp resulted, from which clones were generated. Of the 32 clones that were screened, the largest (C19) included the translation initiation codon and 56 bases of 5′ untranslated region, and overlapped at its 3′ end with the sequence of clone A1. The strategy for RT-PCR and RACE is summarized in Figure 4 (upper panel).

We determined the sequence of the cytosolic β-glucosidase cDNA by directly sequencing overlapping PCR products. Using a series of primers in both the sense and antisense orientations, we sequenced the full length of both strands of the coding region of the cDNA (Figure 4, lower panel). Each segment of each strand was sequenced at least twice, using as template separate PCR products amplified from first-strand cDNA. The complete sequence is shown in Figure 5. It consists of 1671 nucleotides from the 5′-untranslated region to the start of the poly(A) tail. The open reading frame codes for a total of 469 amino acids; the predicted mass of 53744 Da is identical with that estimated by SDS/PAGE (Figure 1). The amino acid composition predicted from the cDNA sequence agrees with that obtained by acid hydrolysis of the purified enzyme (Table 2). The complete sequences of the eight pure tryptic peptides are contained in the coding sequence with only a single discrepancy; the aspartic acid residue at position 16 of peptide 117 (Figure 2, lower panel) is predicted to be the asparagine residue at position 233 in the cDNA sequence (Figure 5). Re-examination of the chromatogram corresponding to the 16th amino acid sequencing cycle for peptide 117 showed the presence of approximately equimolar concentrations of the phenylthiohydantoin derivatives of aspartic acid and asparagine. On the basis of this evidence, we concluded that significant hydrolysis of the amide bond of the side chain of Asn-233 had occurred, and that the correct residue at this position is asparagine. Fractions 92 and 50 of the tryptic digest yielded two or more amino acids at each sequencing cycle. This indicated that two or more peptides were present in fractions 92 and 50. By comparison of the amino acids resulting from each sequencing cycle with the amino acid sequence predicted from the cDNA, the sequences of three peptides from fraction 92 and one peptide from fraction 50 were deduced (Figure 2, lower panel).

The amino acid sequence contains potential sites for post-translational glycosylation and phosphorylation. A consensus sequence for N-glycosylation (NQT) was observed at Asn-46; however, amino acid sequencing of peptide 89, which began with Asn-46, gave no indication of modification of this residue. Several consensus sequences for phosphorylation by the serine/threonine kinases CK2 [25] and calcium-dependent PKC [26] are present in the sequence: nine for CK2 and seven for PKC. Potential phosphorylation sites in the amino acid sequence are shown in Figure 5. Incubation of purified cytosolic β-glucosidase with [γ-^{32}P]ATP and either PKC or CK2 did not result in the incorporation of detectable [^{32}P]phosphate into the β-glucosidase (results not shown).

Sequence analysis

According to the classification scheme described by Henrissat and Bairoch [27], the cytosolic β-glucosidase is a Family 1 β-glycosidase. These authors classified 482 glycosidase enzymes into 45 families based upon amino acid sequence similarities. Family 1 includes β-glycosidases from prokaryotes, plants and mammals. The amino acid sequences of 18 Family 1 enzymes were compared and aligned with that of the cytosolic β-glucosidase (Figure 6). The two functional β-glycosidase domains of mammalian LPH showed the greatest sequence similarity with the cytosolic β-glucosidase, with domain III of LPH showing a slightly greater similarity to the cytosolic enzyme than domain IV. No homology was observed between the cytosolic β-glucosidase and glucocerebrosidase (Table 3).

Table 3 Homologies of selected Family 1 β-glucosidases and glucocerebrosidase with cytosolic β-glucosidase from guinea pig liver

The full names, GenBank or SwissProt accession numbers, and EC designations are indicated in the legend for Figure 6.

Enzyme source	% Identity	% Similarity
E. coli	27.1	50.1
E. chrysanth.	29.8	58.2
B. napus	34.7	56.2
S. alba	37.1	58.0
M. esculenta	38.0	62.7
T. repens	39.9	62.2
A. sativa	36.9	62.3
LPHIII	45.9	67.5
LPHIV	45.4	66.2
B. circulans	39.4	61.8
C. thermocel.	38.3	61.7
T. maritima	41.1	63.0
C. sacch.	35.7	59.4
B. polymyxa	38.1	61.0
A. faecalis	34.8	58.0
M. bispora	36.4	57.8
C. fimi	33.0	53.9
L. lactis	34.5	55.6
Glucocerebrosidase	14.5	43.5

DISCUSSION

The principal result of this study is the elucidation of the cDNA and amino acid sequences of the cytosolic β-glucosidase from guinea pig liver. Homologies with the amino acid sequences of other β-glucosidases permit the identification of critical active-site residues and the prediction of the general folding pattern and active-site topology. These data also suggest a probable evolutionary lineage for this enzyme.

An inherent problem in sequencing clones derived from PCR products is the high error rate in nucleotide incorporation. This results from the lack of proofreading activity in *Taq* polymerase [16]. One means of avoiding this pitfall is to directly sequence PCR products [24]. To further minimize the possibility of sequence errors introduced by *Taq* polymerase, we sequenced each region of the cDNA at least twice with the products of separate PCR reactions. This included sequencing the full length of both the sense and antisense strands, using multiple sequencing primers to generate overlapping stretches of sequence. The identity of the cDNA sequence was confirmed in two ways. First, the coding sequences of all of the tryptic peptides were represented in the cDNA open reading frame. Secondly, within the margin of error of amino acid composition analysis, the amino acid composition of the pure enzyme matches that predicted from the cDNA sequence (Table 2). The distribution of the peptide sequences along the length of the cDNA sequence reduces the likelihood that the cDNA sequence represents a different but closely related gene product (Figure 5). A total of 142 amino

acids in the cDNA sequence (30%) were represented in non-overlapping peptide fragments.

The consensus sequences for serine/threonine phosphorylation raised the possibility that the cytosolic β-glucosidase might be regulated in response to intracellular second messengers. We can exclude neither the possibility of non-exchangeable phosphates nor the possible requirement for unique *in vivo* conditions for phosphorylation by PKC or CK2. However, in the absence of other evidence, our inability to demonstrate *in vitro* phosphorylation of the enzyme leads us to infer that the phosphorylation consensus sequences for PKC and CK2 are probably cryptic in the native β-glucosidase. Likewise, the absence of heterogeneity seen at Asn-46 during amino acid sequencing and the absence of a signal sequence confirm earlier reports that the cytosolic β-glucosidase is not glycosylated [8–10].

The assignment of the cytosolic β-glucosidase to glycosyl hydrolase Family 1 explains several of its distinguishing catalytic features. First, the cytosolic β-glucosidase will hydrolyse glycopyranosides of either epimer at the C-4 position of the sugar (i.e. β-D-Glc or β-D-Gal). It also hydrolyses glycopyranosides lacking the C-6 hydroxyl (e.g. β-D-Fuc) [7–9]. This broad substrate specificity distinguishes the cytosolic β-glucosidase from glucocerebrosidase [9]. Such a broad substrate specificity is characteristic of Family 1 β-glycosidases [28–30]. Secondly, the cytosolic β-glucosidase shows greater amino acid sequence similarity with domain III of LPH versus domain IV. This supports the inference of Mantei et al. [23] that the glycosylceramidase domain IV of LPH is the phylogenetic precursor of the lactase domain. The guinea pig liver β-glucosidase and domain III of LPH share several catalytic properties, including the ability to hydrolyse β(1-6) disaccharides [11,31] and poor activity towards glucosylceramide [9,30]. The cytosolic β-glucosidase may have evolved from a partial duplication of the LPH gene, similar to the duplications which resulted in the contemporary structure of LPH [23].

Recent studies of other Family 1 β-glycosidases have provided structural and mechanistic insights directly applicable to the cytosolic β-glucosidase. Withers and associates, studying the β-glucosidase from *Agrobacterium faecalis* (EC 3.2.1.21), have conclusively identified Glu-358 as the catalytic nucleophile [32]. This residue lies in the highly conserved consensus sequence YITENG, and it forms the covalent enzyme sugar intermediate which is characteristic of configuration-retaining glycosidases [33]. Using covalent labelling methods, Keresztessy et al., have shown that Glu-198 of the β-glucosidase from *Manihot esculenta* (EC 3.2.1.21) is the proton donor responsible for catalysing the glycosylation and deglycosylation steps in the catalytic mechanism [34]. This residue lies in another consensus sequence (INEP) that is highly conserved among Family 1 glycosidases. The X-ray crystal structures of the Family 1 β-glycosidases from *Lactococcus lactis* (EC 3.2.1.85) and *Trifolium repens* (EC 3.2.1.21) confirm the identity of the proton donor [35,36]. The crystal structures also demonstrate that Family 1 β-glycosidases share a common $(\alpha/\beta)_8$ folding motif with the active site located in a pocket lined with charged and aromatic residues [35,36]. The amino acid residues corresponding to the catalytic nucleophile and the acid catalyst in the cytosolic β-glucosidase are residues Glu-373 and Glu-165 respectively (Figures 5 and 6).

In summary, the cDNA sequence of the cytosolic β-glucosidase establishes the evolutionary lineage for this enzyme, and accounts for many of the enzyme's previously reported catalytic properties. These data provide the foundation for future studies of the structure, function and regulation of expression of this enzyme.

We acknowledge the assistance of the UNM Health Sciences Center Protein Chemistry Lab in the synthesis of oligonucleotides and for automated peptide sequencing. We also thank Drew Kalishman for assistance with protein purification. This work was supported by a grant from the Research Allocation Committee of the UNM Health Sciences Center.

REFERENCES

1 Auricchio, S., Rubino, A., Landolt, M., Semenza, G. and Prader, A. (1963) Lancet **2**, 324–326
2 Dahlqvist, A., Hammond, J. D., Crane, R. K., Dunphy, J. V. and Littman, A. (1963) Gastroenterology **45**, 488–491
3 Semenza, G. and Auricchio, S. (1989) in The Metabolic Basis of Inherited Disease, vol. 2., (Scriver, C. R., Beaudet, A. L., Sly, W. S. and Valle, D., eds.), pp. 2975–2997, McGraw-Hill, New York
4 Brady, R. O., Kanfer, J. N. and Shapiro, D. (1965) Biochem. Biophys. Res. Commun. **18**, 221–225
5 Patrick, A. D. (1965) Biochem. J. **97**, 17c–18c
6 Glew, R. H., Gopalan, V., Forsyth, G. W. and Vander Jagt, D. J. (1993) in β-Glucosidases: Biochemistry and Molecular Biology, pp. 83–112, American Chemical Society, Washington, DC
7 Abraham, H. E. and Robinson, D. (1969) Biochem. J. **111**, 749–755
8 Glew, R. H., Peters, S. P. and Christopher, A. R. (1976) Biochim. Biophys. Acta **422**, 179–199
9 Daniels, L. B., Coyle, P. J., Chiao, Y. B., Glew, R. H. and Labow, R. S. (1981) J. Biol. Chem. **256**, 13004–13013
10 LaMarco, K. L. and Glew, R. H. (1986) Biochem. J. **237**, 469–476
11 Gopalan, V., Pastuszyn, A. P., Galey, W. R. and Glew, R. H. (1992) J. Biol. Chem. **267**, 14027–14032
12 Gopalan, V., Vander Jagt, D. J., Libell, D. P. and Glew, R. H. (1992) J. Biol. Chem. **267**, 9629–9638
13 Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254
14 Cohen, S. A. and Strydom, D. J. (1988) Anal. Biochem. **174**, 1–16
15 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
16 Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharff, J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) Science **239**, 487–491
17 Yamada, T., Wheeler, C. M., Halpern, A. L., Stewart, A.-C. M. and Jenison, S. A. (1995) J. Virol. **69**, 7743–7753
18 Tabor, S. and Richardson, C. C. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 4767–4771
19 Frohman, M. (1993) Methods Enzymol. **218**, 340–356
20 Hultman, T., Stahl, S., Holmes, E. and Uhlen, M. (1989) Nucleic Acids Res. **17**, 4937–4946
21 Orr, A., Ivanova, V. S. and Bonner, W. M. (1995) Biotechniques **19**, 149–151
22 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
23 Mantei, N., Villa, M., Enzler, T., Wacker, H., Boll, W., James, P., Hunziker, W. and Semenza, G. (1988) EMBO J. **7**, 2705–2713
24 Rao, V. B. (1994) Anal. Biochem. **216**, 1–14
25 Pinna, L. A. (1990) Biochim. Biophys. Acta **1054**, 267–284
26 Woodget, J. R., Gould, K. L. and Hunter, T. (1986) Eur. J. Biochem. **161**, 177–184
27 Henrissat, B. and Bairoch, A. (1993) Biochem. J. **293**, 781–788
28 Day, A. G. and Withers, S. G. (1986) Biochem. Cell Biol. **64**, 914–922
29 Yeoh, H.-K. and Woo, H.-C. (1992) Phytochemistry **31**, 2263–2265
30 Colombo, V., Lorenz-Meyer, H. and Semenza, G. (1973) Biochim. Biophys. Acta **327**, 412–424
31 Freund, J.-N., Gosse, F. and Raul, F. (1991) Enzyme **45**, 71–74
32 Withers, S. G., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B. and Aebersold, R. (1990) J. Am. Chem. Soc. **112**, 5887–5889
33 McCarter, J. D. and Withers, S. G. (1994) Curr. Opin. Struct. Biol. **4**, 885–892
34 Keresztessy, Z., Kiss, L. and Hughes, M. A. (1994) Arch. Biochem. Biophys. **315**, 323–330
35 Weisman, C., Beste, G., Hengstenberg, W. and Schulz, G. (1995) Structure **3**, 961–968
36 Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J. and Hughes, M. A. (1995) Structure **3**, 951–960

Received 21 March 1996/7 June 1996; accepted 25 June 1996