Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 685–689, February 1987
Biochemistry

Cloning, sequencing, and expression of cDNA for human β-glucuronidase

(lysosomal enzymes/mucopolysaccharidosis type VII/COS cells/simian virus 40/alternate splicing)

AKIHIRO OSHIMA*, JOHN W. KYLE*, RAYMOND D. MILLER*, JOSEPH W. HOFFMANN*, PENELOPE P. POWELL*, JEFFREY H. GRUBB*, WILLIAM S. SLY*†, MICHAEL TROPAK‡, KEVIN S. GUISE‡, AND ROY A. GRAVEL‡§

*E. A. Doisy Department of Biochemistry, St. Louis University School of Medicine, St. Louis, MO 63104; ‡Research Institute, Hospital for Sick Children, Toronto, ON, Canada M5G 1X8; and §Department of Medical Genetics, University of Toronto, Toronto, ON, Canada M5S 1A8

Communicated by Stuart Kornfeld, October 6, 1986 (received for review August 1, 1986)

ABSTRACT **We report here the cDNA sequence for human placental β-glucuronidase (β-D-glucuronoside glucuronosohydrolase, EC 3.2.1.31) and demonstrate expression of the human enzyme in transfected COS cells. We also sequenced a partial cDNA clone from human fibroblasts that contained a 153-base-pair deletion within the coding sequence and found a second type of cDNA clone from placenta that contained the same deletion. Nuclease S1 mapping studies demonstrated two types of mRNAs in human placenta that corresponded to the two types of cDNA clones isolated. The NH_2-terminal amino acid sequence determined for human spleen β-glucuronidase agreed with that inferred from the DNA sequence of the two placental clones, beginning at amino acid 23, suggesting a cleaved signal sequence of 22 amino acids. When transfected into COS cells, plasmids containing either placental clone expressed an immunoprecipitable protein that contained N-linked oligosaccharides as evidenced by sensitivity to endoglycosidase F. However, only transfection with the clone containing the 153-base-pair segment led to expression of human β-glucuronidase activity. These studies provide the sequence for the full-length cDNA for human β-glucuronidase, demonstrate the existence of two populations of mRNA for β-glucuronidase in human placenta, only one of which specifies a catalytically active enzyme, and illustrate the importance of expression studies in verifying that a cDNA is functionally full-length.**

The acid hydrolase β-glucuronidase (β-D-glucuronoside glucuronosohydrolase, EC 3.2.1.31) is a tetrameric glycoprotein composed of identical subunits (M_r = 75,000). It acts as an exoglycosidase in lysosomes, cleaving β-glucuronosyl residues at the nonreducing end of oligosaccharides from glycosaminoglycans. Genetic deficiency of this enzyme leads to accumulation of undegraded glycosaminoglycans in lysosomes and produces the clinical disorder mucopolysaccharidosis type VII (1).

Studies of human β-glucuronidase were important in the discovery of the mannose 6-phosphate recognition marker, which targets acid hydrolases to lysosomes (2, 3), in the delineation of the mannose 6-phosphate receptor-dependent pathways for segregation of acid hydrolases to lysosomes (4, 5) and in animal model studies of enzyme replacement therapy for lysosomal storage diseases (6–8). In this communication we report the nucleotide sequence for a full-length cDNA clone for human β-glucuronidase from human placenta and present data demonstrating expression of human β-glucuronidase in transfected COS-7 cells.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Materials. Human spleen β-glucuronidase (3) and goat antibody raised to this enzyme (9) were prepared as reported. ^{32}P-labeled nucleotides and [^{35}S]methionine (1127 Ci/mmol; 1 Ci = 37 GBq) were from New England Nuclear, Amersham, and ICN. Restriction enzymes and enzymes used in cloning, nuclease S1 protection studies, and DNA sequencing were from New England Biolabs and Bethesda Research Laboratories. Endoglycosidase F (Endo F) (grade II) was from Boehringer Mannheim. COS-7 cells were the gift of M. Green (Institute for Molecular Virology, St. Louis University Medical School, St. Louis, MO). pJC119 was provided by R. A. Lazzarini (Laboratory of Molecular Genetics, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health). IgGsorb was from The Enzyme Center (Malden, MA). EN^{3}HANCE was from New England Nuclear.

Methods. A phage λgt11 cDNA library made from human placental mRNA was provided by J. Evan Sadler (Howard Hughes Medical Institute, Washington University School of Medicine, St. Louis, MO). The cDNA library was screened by plaque hybridization (10). Probes (as described in *Results*) were labeled with ^{32}P by nick-translation. Positive clones were plaque-purified and subcloned into a plasmid derived from pBR325. pHUGP13 and pHUGP15 were sequenced by the dideoxy chain-termination method of Sanger *et al.* (11) with the phage M13 universal sequencing primer or with synthetic oligonucleotide primers (12) after they had been subcloned into phage M13 vectors (13). pHUGF was sequenced by a modification of the Sanger dideoxy chain-termination method (14) that involved sequencing double-stranded deletion fragments that had been generated by exonuclease III (15) or BAL-31 (10) and subcloned into vectors SP64 or SP65 (16). Nuclease S1 mapping was performed as outlined by Maniatis *et al.* (10).

Transfection of COS-7 Cells. The cDNA inserts were isolated by agarose gel electrophoresis after a partial *Sal* I digestion and were ligated into the *Sal* I-compatible, unique *Xho* I site of pJC119, which expresses cloned inserts using the simian virus 40 (SV40) late promoter (17). A DEAE-dextran procedure (18) followed by treatment with 100 μM chloroquine (19) was used to transfect COS-7 cells (20). β-Glucuronidase activity in cell extracts and media was determined fluorometrically with 4-methylumbelliferyl-β-glucuronide (21).

Labeling and Immunoprecipitation. At 48 hr posttransfection COS-7 cells were washed, incubated for 2 hr in methionine-free Dulbecco's modified Eagle's medium (DMEM), and labeled for 1 hr with 125 μCi of [^{35}S]methionine (1127

Abbreviations: Endo F, endoglycosidase F; SV40, simian virus 40.
†To whom reprint requests should be addressed.

Ci/mmol) in 1 ml of methionine-free DMEM containing 5% dialyzed fetal bovine serum. After the pulse the cells were washed twice, and normal medium supplemented with 2 mM methionine was added for 3 hr. The cells were harvested, pelleted, and solubilized in 0.2 ml of solubilization buffer (10 mM Tris·HCl, pH 8.5/1 mM $MgCl_2$/0.5% sodium deoxycholate/1% Nonidet P-40/10 μg of aprotinin per ml/0.1% $NaDodSO_4$). The [^{35}S]methionine-labeled extracts were precleared by addition of 5 μl of preimmune goat serum, followed by addition of 100 μl of 20% IgGsorb, incubation at 4°C for 1 hr, and centrifugation. Goat anti-human β-glucuronidase (10 μl) was added to each supernatant, and the samples were incubated for 6 hr at 4°C. The immunoprecipitates were collected after the addition of 100 μl of 20% IgGsorb, incubation at 4°C for 1 hr, and centrifugation for 1 min in a Fisher microcentrifuge. The pellets were washed five times with a buffer containing 1% sodium deoxycholate, 0.4% Nonidet P-40, 0.5 M NaCl, and 150 mM Tris·HCl (pH 7.4), washed once with 10 mM Tris·HCl (pH 7.4), and then analyzed by $NaDodSO_4$/polyacrylamide gel electrophoresis (22). After electrophoresis, the gels were fixed and stained with Coomassie blue, destained, and treated with EN-^{3}HANCE. The gels were then dried on a gel dryer, and the radioactivity was detected on Kodak X-Omat R film.

Amino Acid Sequence of β-Glucuronidase. The NH_2-terminal amino acid sequence was determined by Edman degradation using polyacrylamide gel-purified human spleen β-glucuronidase (23). The sequence analysis was performed by the Protein Chemistry Facility of Washington University School of Medicine (St. Louis, MO), using an Applied Biosystems vapor-phase sequenator.

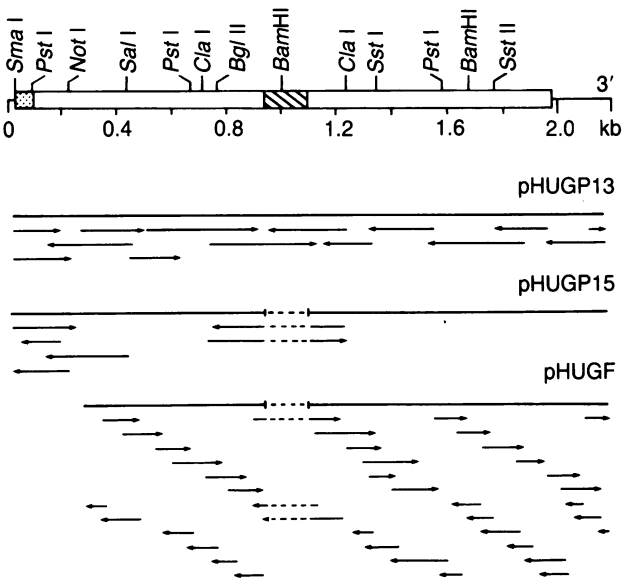

FIG. 1. Restriction map and sequencing strategy for human β-glucuronidase cDNA clones. The composite restriction map is shown at the top of the figure. The large bar represents the entire coding region of the gene; the stippled portion, the region coding for the signal sequence; the hatched portion, the 153-bp sequence found in pHUGP13 but not in pHUGP15 or pHUGF. The three cDNA inserts are aligned, and their linker and poly(A) tail regions are not shown. Arrows below each cDNA insert show the relative position, extent, and direction of DNA sequence determinations.

RESULTS

Isolation of Placental cDNA Clones Encoding β-Glucuronidase. Of 8×10^6 plaques, 40 hybridized to a 0.4-kilobase (kb) probe from the 5′ end of pHUGF (Fig. 1; referred to as pHUG87 in ref. 24). Eight contained inserts that were >2 kb in length. Restriction maps placed these into two different

```
                                                Met Ala Arg Gly Ser Ala Val Ala Trp Ala Ala Leu Gly Pro Leu Leu Trp Gly Cys Ala Leu Gly↓Leu Gln Gly Gly Met Leu Tyr   29
GG  TGG CCG AGC GGG GGA CCG GGA AGC ATG GCC CGG GGG TCG GCG GTT GCC TGG GCG GCG CTC GGG CCG TTG TTG TGG GGC TGC GCG CTG GGG CTG CAG GGC GGG ATG CTG TAC  113
Pro Gln Glu Ser Pro Ser Arg Glu Cys Lys Glu Leu Asp Gly Leu Trp Ser Phe Arg Ala Asp Phe Ser Asp Asn Arg Arg Arg Gly Phe Glu Glu Gln Trp Tyr Arg Arg Pro   67
CCC CAG GAG AGC CCG TCG CGG GAG TGC AAG GAG CTG GAC GGC CTC TGG AGC TTC CGC GCC GAC TTC TCT GAC AAC CGA CGC CGG GGC TTC GAG GAG CAG TGG TAC CGG CGG CCG  227
Leu Trp Glu Ser Gly Pro Thr Val Asp Met Pro Val Pro Ser Ser Phe Asn Asp Ile Ser Gln Asp Trp Arg Leu Arg His Phe Val Gly Trp Val Trp Tyr Glu Arg Glu Val  105
CTG TGG GAG TCA GGC CCC ACC GTG GAC ATG CCA GTT CCC TCC AGC TTC AAT GAC ATC AGC CAG GAC TGG CGT CTG CGG CAT TTT GTC GGC TGG GTG TGG TAC GAA CGG GAG GTG  341
Ile Leu Pro Glu Arg Trp Thr Gln Asp Leu Arg Thr Arg Val Val Leu Arg Ile Gly Ser Ala His Ser Tyr Ala Ile Val Trp Val Asn Gly Val Asp Thr Leu Glu His Glu  143
ATC CTG CCG GAG CGA TGG ACC CAG GAC CTG CGC ACA AGA GTG GTG CTG AGG ATT GGC AGT GCC CAT TCC TAT GCC ATC GTG TGG GTG AAT GGG GTC GAC ACG CTA GAG CAT GAG  455
Gly Gly Tyr Leu Pro Phe Glu Ala Asp Ile Ser Asn Leu Val Gln Val Gly Pro Leu Pro Ser Arg Leu Arg Ile Thr Ile Ala Ile Asn Asn Thr Leu Thr Pro Thr Thr Leu  181
GGG GGC TAC CTC CCC TTC GAG GCC GAC ATC AGC AAC CTG GTC CAG GTG GGG CCC CTG CCC TCC CGG CTC CGA ATC ACT ATC GCC ATC AAC AAC ACA CTC ACC CCC ACC ACC CTG  569
Pro Pro Gly Thr Ile Gln Tyr Leu Thr Asp Thr Ser Lys Tyr Pro Lys Gly Tyr Phe Val Gln Asn Thr Tyr Phe Asp Phe Phe Asn Tyr Ala Gly Leu Gln Arg Ser Val Leu  219
CCA CCA GGG ACC ATC CAA TAC CTG ACT GAC ACC TCC AAG TAT CCC AAG GGT TAC TTT GTC CAG AAC ACA TAT TTT GAC TTT TTC AAC TAC GCT GGA CTG CAG CGG TCT GTA CTT  683
Leu Tyr Thr Thr Pro Thr Thr Tyr Ile Asp Asp Ile Thr Val Thr Thr Ser Val Glu Gln Asp Ser Gly Leu Val Asn Tyr Gln Ile Ser Val Lys Gly Ser Asn Leu Phe Lys  257
CTG TAC ACG ACA CCC ACC ACC TAC ATC GAT GAC ATC ACC GTC ACC ACC AGC GTG GAG CAA GAC AGT GGG CTG GTG AAT TAC CAG ATC TCT GTC AAG GGC AGT AAC CTG TTC AAG  797
Leu Glu Val Arg Leu Leu Asp Ala Glu Asn Lys Val Val Ala Asn Gly Thr Gly Thr Gln Gly Gln Leu Lys Val Pro Gly Val Ser Leu Trp Trp Pro Tyr Leu Met His Glu  295
TTG GAA GTG CGT CTT TTG GAT GCA GAA AAC AAA GTC GTG GCG AAT GGG ACT GGG ACC CAG GGC CAA CTT AAG GTG CCA GGT GTC AGC CTC TGG TGG CCG TAC CTG ATG CAC GAA  911
Arg Pro Ala Tyr Leu Tyr Ser Leu Glu Val Gln Leu Thr Ala Gln Thr Ser Leu Gly Pro Val Ser Asp Phe Tyr Thr Leu Pro Val Gly Ile Arg Thr Val Ala Val Thr Lys  333
CGC CCT GCC TAT CTG TAT TCA TTG GAG GTG CAG CTG ACT GCA CAG ACG TCA CTG GGG CCT GTG TCT GAC TTC TAC ACA CTC CCT GTG GGG ATC CGC ACT GTG GCT GTC ACC AAG 1025
Ser Gln Phe Leu Ile Asn Gly Lys Pro Phe Tyr Phe His Gly Val Asn Lys His Glu Asp Ala Asp Ile Arg Gly Lys Gly Phe Asp Trp Pro Leu Leu Val Lys Asp Phe Asn  371
AGC CAG TTC CTC ATC AAT GGG AAA CCT TTC TAT TTC CAC GGT GTC AAC AAG CAT GAG GAT GCG GAC ATC CGA GGG AAG GGC TTC GAC TGG CCG CTG CTG GTG AAG GAC TTC AAC 1139
Leu Leu Arg Trp Leu Gly Ala Asn Ala Phe Arg Thr Ser His Tyr Pro Tyr Ala Glu Glu Val Met Gln Met Cys Asp Arg Tyr Gly Ile Val Val Ile Asp Glu Cys Pro Gly  409
CTG CTT CGC TGG CTT GGT GCC AAC GCT TTC CGT ACC AGC CAC TAC CCC TAT GCA GAG GAA GTG ATG CAG ATG TGT GAC CGC TAT GGG ATT GTG GTC ATC GAT GAG TGT CCC GGC 1253
Val Gly Leu Ala Leu Pro Gln Phe Phe Asn Asn Val Ser Leu His His His Met Gln Val Met Glu Glu Val Val Arg Arg Asp Lys Asn His Pro Ala Val Val Met Trp Ser  447
GTG GGC CTG GCG CTG CCG CAG TTC TTC AAC AAC GTT TCT CTG CAT CAC CAC ATG CAG GTG ATG GAA GAA GTG GTG CGT AGG GAC AAG AAC CAC CCC GCG GTC GTG ATG TGG TCT 1367
Val Ala Asn Glu Pro Ala Ser His Leu Glu Ser Ala Gly Tyr Tyr Leu Lys Met Val Ile Ala His Thr Lys Ser Leu Asp Pro Ser Arg Pro Val Thr Phe Val Ser Asn Ser  485
GTG GCC AAC GAG CCT GCG TCC CAC CTA GAA TCT GCT GGC TAC TAC TTG AAG ATG GTG ATC GCT CAC ACC AAA TCC TTG GAC CCC TCC CGG CCT GTG ACC TTT GTG AGC AAC TCT 1481
Asn Tyr Ala Ala Asp Lys Gly Ala Pro Tyr Val Asp Val Ile Cys Leu Asn Ser Tyr Tyr Ser Trp Tyr His Asp Tyr Gly His Leu Glu Leu Ile Gln Leu Gln Leu Ala Thr  523
AAC TAT GCA GCA GAC AAG GGG GCT CCG TAT GTG GAT GTG ATC TGT TTG AAC AGC TAC TAC TCT TGG TAT CAC GAC TAC GGG CAC CTG GAG TTG ATT CAG CTG CAG CTG GCC ACC 1595
Gln Phe Glu Asn Trp Tyr Lys Lys Tyr Gln Lys Pro Ile Ile Gln Ser Glu Tyr Gly Ala Glu Thr Ile Ala Gly Phe His Gln Asp Pro Pro Leu Met Phe Thr Glu Glu Tyr  561
CAG TTT GAG AAC TGG TAT AAG AAG TAT CAG AAG CCC ATT ATT CAG AGC GAG TAT GGA GCA GAA ACG ATT GCA GGG TTT CAC CAG GAT CCA CCT CTG ATG TTC ACT GAA GAG TAC 1709
Gln Lys Ser Leu Leu Glu Gln Tyr His Leu Gly Leu Asp Gln Lys Arg Arg Lys Tyr Val Val Gly Glu Leu Ile Trp Asn Phe Ala Asp Phe Met Thr Glu Gln Ser Pro Thr  599
CAG AAA AGT CTG CTA GAG CAG TAC CAT CTG GGT CTG GAT CAA AAA CGC AGA AAA TAT GTG GTT GGA GAG CTC ATT TGG AAT TTT GCC GAT TTC ATG ACT GAA CAG TCA CCG ACG 1823
Arg Val Leu Gly Asn Lys Lys Gly Ile Phe Thr Arg Gln Arg Gln Pro Lys Ser Ala Ala Phe Leu Leu Arg Glu Arg Tyr Trp Lys Ile Ala Asn Glu Thr Arg Tyr Pro His  637
AGA GTG CTG GGG AAT AAA AAG GGG ATC TTC ACT CGG CAG AGA CAA CCA AAA AGT GCA GCG TTC CTT TTG CGA GAG AGA TAC TGG AAG ATT GCC AAT GAA ACC AGG TAT CCC CAC 1937
Ser Val Ala Lys Ser Gln Cys Leu Glu Asn Ser Pro Phe Thr *                                                                                                  651
TCA GTA GCC AAG TCA CAA TGT TTG GAA AAC AGC CCG TTT ACT TGA GCAAGACTGATACCACCTGCGTGTCCCTTCCTCCCCGAGTCAGGGCGACTTCCACAGCAGCAGAACAAGTGCCTCCTGGACTGTTCACGGCAGACCAG 2080

AACGTTTCTGGCCTGGGTTTTGTGGTCATCTATTCTAGCAGGGAACACTAAAGGTGGAAATAAAAGATTTTCTATTATGGAAATAAAGAGTTGGCATGAAAGTCGCTACTG 2191
```

FIG. 2. Nucleotide and deduced amino acid sequence of human β-glucuronidase. The nucleotides are numbered in the 5′-to-3′ direction, starting with the first nucleotide of the insert in pHUGP13. The amino acids are numbered beginning with the first residue of the signal sequence. The proposed signal sequence cleavage site is indicated by an arrow. The 10 amino acids that correspond to the NH_2-terminal amino acids of human spleen β-glucuronidase are indicated by the overline. The four potential N-linked glycosylation sites are boxed. The 153 nucleotides not present in pHUGP15 and pHUGF are bracketed and underlined. The potential polyadenylylation signals (AATAAA) in the 3′ untranslated region are underlined.

groups typified by pHUGP13 and pHUGP15 (Fig. 1), which were both 0.25 kb longer at the 5′ end than pHUGF. The pHUGP13-type clones contain an additional 0.15-kb internal segment not present in pHUGP15 or pHUGF (Fig. 1).

Nucleotide Sequence Analysis. Fig. 2 shows the 2191-nucleotide sequence obtained by sequencing pHUGF, the partial cDNA clone from human fibroblasts, and the two placental cDNA clones; the strategy used is shown in Fig. 1. The 5′ end of pHUGP13 was numbered 1; pHUGP15 began at base pair 8 and pHUGF at base pair 267. The additional internal segment in pHUGP13 contained 153 base pairs (bp) (base pairs 939–1091). The poly(A) tract in pHUGP13 begins at base pair 2180. All other regions sequenced were identical. There were two potential poly(A) addition sites (AATAAA) 25 and 48 bp upstream from the poly(A) tract (Fig. 2).

Nuclease S1 Mapping Studies of Human Placental mRNA. Nuclease S1 mapping studies were carried out to determine whether there are two distinct mRNAs for β-glucuronidase in placenta (Fig. 3). Probes used for the nuclease S1 protection analysis were the 527-nucleotide *Cla* I–*Cla* I fragment from pHUGP13 (nucleotides 710–1237) and the 374-nucleotide *Cla* I–*Cla* I fragment from pHUGF (Fig. 1). Lane A in Fig. 3 shows the protection of bands of approximately 527 and 146 nucleotides by the *Cla* I–*Cla* I fragment from pHUGP13. The band near 527 nucleotides would represent a mRNA that is completely homologous with the probe. The band near 146 nucleotides would represent a mRNA that had diverged at a distance of 146 nucleotides from the 3′ end of the probe. This position would correspond with the 3′ border of the 153-bp segment and would be expected for the pHUGP15 mRNA. Lane C in Fig. 3 shows the results when the 374-nucleotide *Cla* I–*Cla* I fragment from pHUGF was used as a probe. Two S1 nuclease-resistant bands can be seen (approximately 146 and 374 nucleotides). The sizes are those predicted for the probe missing the 153-bp segment. Lanes B and D in Fig. 3 show controls, where each of the probes were hybridized to yeast tRNA. These results provide evidence for two populations of mRNA in placenta that correspond to the two types of cDNA isolated (pHUGP13 and pHUGF).

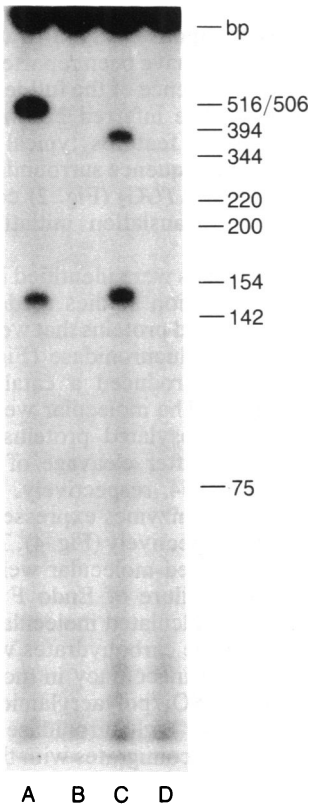

FIG. 3. Autoradiography of protected bands from nuclease S1 mapping of human placental RNA. Approximately 0.1 μg of *Cla* I–*Cla* I cDNA probe labeled at the 5′ end with [γ-^{32}P]ATP was hybridized in a total volume of 30 μl for 3 hr at 58°C to either 200 μg of total human placental RNA (lanes A and C) or yeast tRNA (lanes B and D) (13). After hybridization, nuclease S1 was added to 2000 units per ml, and the 300-μl mixtures were incubated at 40°C for 1 hr. The samples were then precipitated with ethanol after the addition of 20 μg of carrier yeast tRNA and analyzed by gel electrophoresis on a 5% polyacrylamide gel containing 7 M urea. The probe for lanes A and B was the 527-nucleotide *Cla* I–*Cla* I fragment from pHUGP13. The probe for lanes C and D was the 374-nucleotide *Cla* I–*Cla* I fragment from pHUGF.

Predicted Amino Acid Sequence. The nucleotide sequence for the longer pHUGP13 clone codes for a protein of 651 residues, with a calculated M_r of 74,715 (Fig. 2). The initiation methionine codon begins at base pair 27. The first in-frame stop codon (TGA) begins at base pair 1980. The first 10 NH_2-terminal amino acids determined by Edman degradation of human spleen β-glucuronidase corresponded exactly with the deduced amino acid sequence of the human placental cDNA beginning at residue 23 (Fig. 2). After cleavage of the 22-residue signal sequence, the unglycosylated protein encoded by pHUGP13 would contain 629 amino acids and have a M_r of 72,562. The 153-bp deletion in pHUGP15 does not disrupt the open reading frame. Thus, pHUGP15 codes for a protein missing an internal stretch of 51 amino acids that would have M_r values of 69,107 and 66,954 before and after cleavage of the signal sequence.

Several previous studies incorrectly concluded that there were no cysteine residues in β-glucuronidase (9, 25–28). There are actually six cysteine residues in the inferred amino acid sequence (Fig. 2). The amino acid sequences surrounding three of these agree with cysteine-containing tryptic peptides from rat preputial β-glucuronidase (29). The predicted amino acid sequence contains four potential sites for asparagine-linked glycosylation (Fig. 2).

β-Glucuronidase sequence was compared with the 3800 sequences found in the protein database of the National Biomedical Research Foundation as described (30, ¶). The only significant homology (16%) was with the first 668 amino acids of *Escherichia coli* β-galactosidase. The region from amino acid 324 to amino acid 479 of β-glucuronidase had the highest homology (32%). No significant protein homology was detected with the same program when comparing β-glucuronidase with other sequenced lysosomal enzymes, including β-hexosaminidase (31, 32), glucocerebrosidase (33), α-fucosidase (34), cathepsin D (35), and α-galactosidase A (36) or with the microsomal enzyme UDP-glucuronosyltransferase (37).

Expression and Immunoprecipitation. Inserts from pHUGP13 and pHUGP15 were subcloned into the SV40 expression vector pJC119 (17) and were designated pSVL(JC)-HUGP13 and pSVL(JC)-HUGP15, respectively. Cells were transfected with each of these two clones or with pJC119 and labeled with [^{35}S]methionine as described. The ^{35}S-labeled proteins were immunoprecipitated by antibody to human β-glucuronidase and analyzed by gel electrophoresis. Transfection with plasmid containing either insert led to a large increase in immunoprecipitable protein (Fig. 4, lanes 3 and 4) compared with cells transfected with vector alone (Fig. 4, lane 2). Normal human fibroblasts labeled and immunoprecipitated under identical conditions are shown in lane 1 of Fig. 4, where faint bands at M_r 82,000 and M_r 79,000 are seen. pSVL(JC)-HUGP13 produced a protein of M_r 82,000. pSVL(JC)-HUGP15 produced a labeled immunoprecipitable

¶Protein Identification Resource (1986) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 10.0.

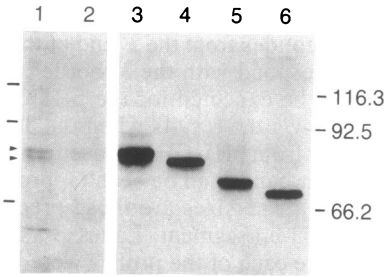

FIG. 4. Fluorography of [^{35}S]methionine-labeled proteins expressed after transfection of COS-7 cells. At 48 hr posttransfection, COS-7 cells and nontransfected fibroblasts were labeled with [^{35}S]-methionine and immunoprecipitated with antiserum to human β-glucuronidase as described. Proteins immunoprecipitated from lysed extracts were eluted from the IgGsorb pellets with 40 μl of 1% $NaDodSO_4$ in 50 mM Tris·HCl (pH 6.8) by boiling for 5 min. The eluted [^{35}S]methionine-labeled protein was split into two aliquots, one of which was treated with 1 milliunit of Endo F in 50 mM citrate buffer (pH 5.5) for 20 hr. Both aliquots were then analyzed by $NaDodSO_4$/polyacrylamide gel electrophoresis. Lanes: 1, immunoprecipitated protein from normal human fibroblasts; 2, COS-7 cells transfected with vector (pJC119) alone; 3, immunoprecipitate from cells transfected with pSVL(JC)-HUGP13; 4, immunoprecipitate from cells transfected with pSVL(JC)-HUGP15; 5, Endo F-treated sample from lane 3; 6, Endo F-treated sample from lane 4. The arrowheads at the left margin correspond to the mobilities of the proteins in lanes 3 and 4. All lanes were from the same gel exposed for either 5 hr (lanes 3–6) or 12 hr (lanes 1 and 2).

protein of M_r 79,000. The antibody to human β-glucuronidase precipitates the endogenous COS cell enzyme. A faint endogenous β-glucuronidase band was observed in COS cells transfected with vector alone (Fig. 4, lane 2). The labeled immunoprecipitable products of transfection with either pSVL(JC)-HUGP13 or pSVL(JC)-HUGP15 contained N-linked carbohydrates as evidenced by sensitivity to Endo F (Fig. 4, lanes 5 and 6). The molecular weights of the Endo F-treated proteins were approximately M_r 74,000 and M_r 71,500 after transfection with pSVL(JC)-HUGP13 and pSVL(JC)-HUGP15, respectively. Both produced a labeled immunoprecipitable protein that was secreted into the media, although the rate of secretion of the protein from transfection with pSVL(JC)-HUGP15 was significantly slower (data not shown).

Expression of β-Glucuronidase Activity. Results from experiments after transfection of COS-7 cells with plasmids containing either of the two types of placental cDNA inserts, or with no insert, are presented in Table 1. Human β-glucuronidase was distinguished from the endogenous COS cell enzyme by its resistance to heat inactivation (7, 8). Transfection with pSVL(JC)-HUGP13 led to an 11-fold increase in intracellular β-glucuronidase activity. However, neither the transfection with plasmid without insert nor the pSVL(JC)-HUGP15 plasmid led to any increase in β-glucuronidase activity. Heat inactivation of the endogenous COS cell enzyme made the difference even more dramatic. The large increase in β-glucuronidase activity in the medium was also due to heat-stable human enzyme (Table 1).

To verify that the lack of expression of enzymatic activity by the pHUGP15 clones was due to the absence of the 153-bp segment, the *Bgl* II–*Sst* I fragment from pHUGP15 (Fig. 1) was replaced by the *Bgl* II–*Sst* I fragment (base pairs 767–1780) from pHUGP13. Transfection of this chimeric cDNA into COS-7 cells resulted in expression of an immunoreactive protein (M_r = 82,000) with heat-stable human β-glucuronidase activity. Replacing the *Bgl* II–*Sst* I fragment in pHUGP13 with that from pHUGF, which lacks the 153-bp segment, produced a chimeric cDNA that expressed a protein (M_r = 79,000) that lacked enzymatic activity. These experiments (data not shown) localized the basis for the loss of enzymatic activity to the deletion of the 153-bp segment.

Table 1. Expression of human β-glucuronidase in transfected COS-7 cells

Plasmid	Cell extract		Medium	
	Total units	Heat resistant	Total units	Heat resistant
pJC119	53	1	10	1
pSVL(JC)-HUGP13	570	484	162	130
pSVL(JC)-HUGP15	52	1	10	2

COS-7 cells were transfected in 60-mm dishes with 10 μg of the indicated plasmid. Plasmid pJC119 is the expression vector with no cDNA insert. At 60 hr posttransfection, the cells and the media were collected as described. Endogenous COS-7 β-glucuronidase was inactivated by heating at 65°C for 90 min after dilution of the extract with an equal volume of 40 mM Tris·HCl, pH 7.5/0.15 M NaCl/10 mg of bovine serum albumin per ml (7, 8). For fluorometric assay of β-glucuronidase, dilutions of heat-treated or untreated extracts and medium were added to 100 μl of assay buffer (10 mM 4-methylumbelliferyl-β-glucuronide/0.1 M acetate buffer, pH 4.8/1 mg of human serum albumin per ml) and incubated at 37°C for 2 hr (cell-lysate) or 10 hr (medium) (21). One unit of enzyme activity was defined as the activity that released 1 nmol of 4-methylumbelliferone per hr.

DISCUSSION

Partial cDNA clones for mouse (38–40), rat (41, 42), and human (24) β-glucuronidase have been reported. In this study we report the complete sequence of the full-length cDNA for human β-glucuronidase. The inferred 22-amino-acid signal sequence has a number of features typical of eukaryotic signal sequences (42). The sequence surrounding the inferred ATG initiator codon (AGC*ATG*G) (Fig. 2) conforms to the consensus sequence for translation initiation sites (AC-C*ATG*G) (43).

Two types of cDNA clones were identified in the placental cDNA library. The expression studies made it clear that, although both clones specified proteins that were precipitated by antiserum to human β-glucuronidase (Fig. 4), only the longer clone, pHUGP13, produced a catalytically active β-glucuronidase (Table 1). The molecular weight calculated for the monomeric unglycosylated proteins produced by pHUGP13 and pHUGP15 after cleavage of the signal sequence are 72,562 and 66,954, respectively. The molecular weights determined for the enzymes expressed in COS cells were 82,000 and 79,000, respectively (Fig. 4). Treatment with Endo F reduced the observed molecular weights to 74,000 and 71,500 (Fig. 4). The failure of Endo F to reduce the expressed proteins to the calculated molecular weights may mean that the removal of the carbohydrates was incomplete or may reflect the inherent inaccuracy in molecular weight determinations by $NaDodSO_4$/polyacrylamide gel electrophoresis (44). The labeled β-glucuronidase produced by transfection with pHUGP13 comigrates with the endogenous COS cell protein and the M_r 82,000 fibroblast β-glucuronidase band. It is not yet clear whether the protein of M_r 79,000 seen in fibroblast extracts represents the equivalent of the pHUGP15 protein or a processed form of the M_r 82,000 β-glucuronidase.

The nuclease S1 mapping studies (Fig. 3) show that mRNAs with and without the 153-bp segment are present in placenta. These may represent the products of alternate splicing. Alternate splicing of internal exons leading to "exon skipping" has been reported for a number of different genes (45). We have not yet established whether the 153-bp segment corresponds to an exon in the β-glucuronidase gene. Determination of the flanking sequences in genomic DNA should provide that answer. The shorter clones could also be

products of a different gene. The gene for human β-glucuronidase has been assigned to chromosome 7 (46). Whatever the origin of the shorter clones, the 51 amino acids encoded by the deleted sequence appear to be essential for catalytic activity of human β-glucuronidase in COS cells.

Whether the shorter protein specified by the pHUGP15 sequence has a biological function is presently unclear. Nevertheless, the existence of the pHUGP15-type clone, which specifies an immunoprecipitable product that lacks catalytic activity, emphasizes the importance of expression studies to support any claim that a cloned gene is full-length. This criterion would appear essential, unless the complete amino acid sequence is already known from amino acid sequence data.

Since this paper was communicated, Nishimura *et al.* (47) reported the cloning and sequencing of the cDNA for rat preputial β-glucuronidase and the expression of an immunoprecipitable protein in a cell-free translation system.

The authors gratefully acknowledge Drs. John Chirgwin, Paul Wollenzien, Kay Nolan, Ms. Helen Cochrane, and Mr. David Bachinsky for helpful advice. This work was supported by National Institutes of Health Grants GM34182 and AI22359 and March of Dimes Grant 6-407 to W.S.S. and Medical Research Council of Canada Grant PG-4 to R.A.G.; J.W.K. was supported by National Institutes of Health Multidisciplinary Postdoctoral Training Grant HL 07050.

1. Sly, W. S., Quinton, B. A., McAlister, W. H. & Rimoin, D. L. (1873) *J. Pediatr.* **82,** 249–257.
2. Kaplan, A., Achord, D. T. & Sly, W. S. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 2026–2030.
3. Natowicz, M. R., Chi, M. M.-Y., Lowry, O. H. & Sly, W. S. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4322–4326.
4. Gonzalez-Noriega, A., Grubb, J. H., Talkad, V. & Sly, W. S. (1980) *J. Cell Biol.* **85,** 839–852.
5. Fischer, H. D., Gonzalez-Noriega, A. & Sly, W. S. (1980) *J. Biol. Chem.* **255,** 5069–5074.
6. Achord, D. T., Brot, F. E., Bell, C. E. & Sly, W. S. (1978) *Cell* **15,** 269–278.
7. Frankel, H. A., Glaser, J. H. & Sly, W. S. (1977) *Pediatr. Res.* **11,** 811–816.
8. Achord, D., Brot, F., Gonzalez-Noriega, A., Sly, W. S. & Stahl, P. (1977) *Pediatr. Res.* **11,** 816–822.
9. Brot, F. E., Bell, C. E. & Sly, W. S. (1978) *Biochemistry* **17,** 385–391.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Strauss, E. C., Kobori, J. A., Siu, G. & Hood, L. E. (1986) *Anal. Biochem.* **154,** 353–360.
13. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
14. Korneluk, R. G., Quan, F. & Gravel, R. A. (1985) *Gene* **40,** 317–323.
15. Guo, L., Yang, R. C. A. & Wu, R. (1983) *Nucleic Acids Res.* **11,** 5521–5540.
16. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* **12,** 7035–7056.
17. Sprague, J., Condra, J. H., Arnheiter, H. & Lazzarini, R. A. (1983) *J. Virol.* **45,** 773–781.
18. Lopata, M. A., Cleveland, D. W. & Sollner-Webb, B. (1984) *Nucleic Acids Res.* **12,** 5707–5717.
19. Luthman, H. & Magnusson, G. (1983) *Nucleic Acids Res.* **11,** 1295–1308.
20. Gluzman, Y. (1981) *Cell* **23,** 175–182.
21. Glaser, J. H. & Sly, W. S. (1973) *J. Lab. Clin. Med.* **82,** 969–977.
22. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–684.
23. Hunkapillar, M. W., Lujan, E., Ostrander, F. & Hood, L. E. (1983) *Methods Enzymol.* **91,** 227–236.
24. Guise, K. S., Korneluk, R. G., Waye, J., Lamhonwah, A.-M., Quan, F., Palmer, R., Ganschow, R. E., Sly, W. S. & Gravel, R. A. (1985) *Gene* **34,** 105–110.
25. Himeno, M., Hashiguchi, Y. & Kato, K. (1974) *J. Biochem. (Tokyo)* **76,** 1243–1252.
26. Tomino, S., Paigen, K., Tulsiani, D. R. P. & Touster, O. (1975) *J. Biol. Chem.* **250,** 8503–8509.
27. Tulsiani, D. R. P., Six, H. & Touster, O. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3080–3084.
28. Lusis, A. J. & Paigen, K. (1978) *J. Biol. Chem.* **253,** 7336–7345.
29. Leighton, P. H., Fisher, W. K., Moon, K. E. & Thompson, E. O. P. (1980) *Aust. J. Biol. Sci.* **33,** 513–520.
30. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
31. Myerowitz, R., Piekarz, R., Neufeld, E. F., Shows, T. B. & Suzuki, K. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7830–7834.
32. Korneluk, R. G., Mahuran, D. J., Neote, K., Klavins, M. H., O'Dowd, B. F., Tropak, M., Willard, H. F., Anderson, M. J., Lowden, J. A. & Gravel, R. A. (1986) *J. Biol. Chem.* **261,** 8407–8413.
33. Tsuji, S., Choudary, P. V., Martin, B. M., Winfield, S., Barranger, J. A. & Ginns, E. I. (1986) *J. Biol. Chem.* **261,** 50–53.
34. Fukushima, H., de Wet, J. R. & O'Brien, J. S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1262–1265.
35. Faust, P. L., Kornfeld, S. & Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4910–4914.
36. Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. & Desnick, R. J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4859–4863.
37. Mackenzie, P. I. (1986) *J. Biol. Chem.* **261,** 6119–6125.
38. Catterall, J. F. & Leary, S. L. (1983) *Biochemistry* **22,** 6049–6053.
39. Palmer, R., Gallagher, P. M., Boyko, W. L. & Ganschow, R. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7596–7600.
40. Watson, G., Felder, M., Rabinow, L., Moore, K., Labarca, C., Tietze, C., Molen, G. V., Bracey, L., Brabant, M., Cai, J. & Paigen, K. (1985) *Gene* **36,** 15–25.
41. Hieber, V. C. (1982) *Biochem. Biophys. Res. Commun.* **104,** 1271–1278.
42. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12,** 5145–5164.
43. Kozak, M. (1986) *Cell* **44,** 283–292.
44. Sundelin, J., Melhus, H., Das, S., Eriksson, U., Lind, P., Tragardh, L., Peterson, P. A. & Rask, L. (1985) *J. Biol. Chem.* **260,** 6481–6487.
45. Leff, S. E., Rosenfeld, M. G. & Evans, R. M. (1986) *Annu. Rev. Biochem.* **55,** 1091–1117.
46. Arrendondo-Vega, F. X., Arfin, S. M. & Smith, M. (1984) *Cytogenet. Cell Genet.* **37,** 403–404.
47. Nishimura, Y., Rosenfeld, M. G., Kreibich, G., Gubler, U., Sabatini, D. D., Adesnik, M. & Andy, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7292–7296.