Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 7292–7296, October 1986
Cell Biology

Nucleotide sequence of rat preputial gland β-glucuronidase cDNA and *in vitro* insertion of its encoded polypeptide into microsomal membranes

(cDNA cloning/lysosomal enzymes/biosynthesis/cotranslational processing/β-galactosidase)

YUKIO NISHIMURA*†, MELVIN G. ROSENFELD*, GERT KREIBICH*, UELI GUBLER‡, DAVID D. SABATINI*, MILTON ADESNIK*, AND ROBIN ANDY*

*Department of Cell Biology and The Kaplan Cancer Center, New York University School of Medicine, New York, NY 10016; and ‡Department of Molecular Genetics, Hoffmann–La Roche, Inc., Nutley, NJ 07110

Contributed by David D. Sabatini, June 10, 1986

ABSTRACT **We have selected the rat preputial gland β-glucuronidase as a model protein to study the sorting of newly synthesized lysosomal hydrolases to the lysosome. The complete coding sequence of β-glucuronidase messenger RNA was determined from the sequences of a group of overlapping cDNA clones isolated from preputial gland cDNA libraries. The β-glucuronidase mRNA primary translation product contains 648 amino acids, including an amino-terminal signal sequence of 22 residues. The polypeptide has four potential sites for the addition of asparagine-linked core oligosaccharides. A 376-residue segment of β-glucuronidase shows extensive homology (23% sequence identity) to a portion of *Escherichia coli* β-galactosidase. This homology most likely reflects an evolutionary relationship between the bacterial and eukaryotic enzymes and the conservation of structural features necessary for the glycosidase activity of both proteins. Translation of mRNA synthesized *in vitro* by transcription of a cDNA containing the entire β-glucuronidase coding region yielded a polypeptide that was immunoprecipitated with anti-β-glucuronidase antiserum and had the same electrophoretic mobility as the primary translation product of natural β-glucuronidase mRNA. In the presence of microsomal membranes, the *in vitro*-synthesized β-glucuronidase underwent cotranslational incorporation into the microsomes, as indicated by removal of the signal sequence and the addition of several oligosaccharide chains. The β-glucuronidase cDNA will provide a useful tool to study the mechanism of mannose phosphorylation and other aspects of the sorting of lysosomal enzymes to lysosomes.**

β-Glucuronidase is a lysosomal enzyme that plays an important role in the degradation of dermatan and keratan sulfates (1). In most tissues, this enzyme appears to be localized only in lysosomes, but in mouse liver and kidney it is also present within the endoplasmic reticulum, associated with egasyn, a hydrophobic polypeptide bound to the luminal face of the membrane (2). The richest known source of β-glucuronidase is the preputial gland of female rats, where this enzyme comprises ≈7% of the total protein (3). Rat preputial gland β-glucuronidase is a tetrameric glycoprotein composed of identical subunits with a molecular mass of ≈75 kDa (4).

It has been shown that β-glucuronidase is synthesized on membrane-bound ribosomes (5) and that a transient amino-terminal insertion signal initiates its cotranslational insertion into the endoplasmic reticulum (5, 6). During this process, the polypeptide acquires several oligosaccharide chains (5, 6) that provide potential sites for the addition of phosphate groups to mannose residues. Indeed, three partially phosphorylated oligosaccharide chains have been found associated with the β-glucuronidase of cultured mouse cells (7, 8). β-Glucuronidase is secreted by I-cell disease fibroblasts (9), which are deficient in the enzyme that phosphorylates mannose residues in newly synthesized lysosomal hydrolases (10–14). Furthermore, secretion of β-glucuronidase by normal cultured cells is markedly enhanced by treatment with tunicamycin (5), which prevents the addition of the core oligosaccharide chains where mannose phosphorylation occurs. It appears, therefore, that β-glucuronidase is sorted to lysosomes by a mechanism that involves the mannose 6-phosphate receptor that has been shown to participate in the sorting of other lysosomal hydrolases (for review, see ref. 15).

The identification of the features of lysosomal polypeptides that determine the phosphorylation of their oligosaccharide chains is likely to require the knowledge of the complete primary sequences of at least several of these proteins. This information can be most easily obtained from cDNA clones, which can also be used in transfection experiments to study the synthesis and sorting of the normal polypeptide and of genetically altered variant forms. Recently, partial cDNA clones for β-glucuronidase from rat preputial gland (16, 17), mouse kidney (18, 19), and human fibroblasts (20) have been obtained, but the complete sequence of the protein has not been reported. We describe in this paper the isolation and characterization of a group of overlapping cDNA clones for the preputial gland enzyme, from which the complete primary sequence of the polypeptide was deduced. An intact cDNA was reconstructed from these clones and used to achieve the *in vitro* synthesis of the β-glucuronidase polypeptide and its incorporation into microsomal vesicles, which was accompanied by cotranslational cleavage of the insertion signal and core glycosylation of the polypeptide chain.

MATERIALS AND METHODS

Materials. These were from the following sources: restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, *Escherichia coli* DNA polymerase I, and T4 DNA polymerase, from New England Biolabs; the Klenow fragment of *E. coli* DNA polymerase and S1 nuclease, from Boehringer Mannheim; RNase H and cloned Moloney murine leukemia virus (Mo-MuLV) reverse transcriptase, from Bethesda Research Laboratories; terminal deoxynucleotidyltransferase, calf thymus DNA-derived oligonucleotide primers, and nucleotides, from P-L Biochemicals; avian myeloblastosis reverse transcriptase from New England

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.

Abbreviations: kb, kilobase(s); bp, base pair(s).

†Present address: Department of Physiological Chemistry, Faculty of Pharmaceutical Sciences, Kyushu University, Fukuoka, 812 Japan.

Nuclear; plasmid pSP64, SP6 polymerase, $m^7G(5')ppp(5')G$, and RNasin, from Promega Biotec (Madison, WI).

RNA Isolation and *in Vitro* Translation. Total RNA was prepared from female rat preputial glands (21). Poly(A)$^+$ RNA was isolated (22) and a fraction enriched in translatable β-glucuronidase mRNA was obtained by preparative electrophoresis in a 1.2% agarose gel containing 10 mM methylmercury hydroxide. Translation of mRNA immunoprecipitation with β-glucuronidase antiserum, and NaDodSO$_4$/polyacrylamide gel electrophoresis analysis were carried out as described (5).

Construction of cDNA Libraries. Oligo(dT)-primed cDNA synthesis was carried out in a 100-μl reaction mixture containing 50 mM Tris·HCl (pH 8.6), 8 mM $MgCl_2$, 50 mM KCl, 2 mM sodium pyrophosphate, 20 mM 2-mercaptoethanol, 0.5 mM dithiothreitol, 1 mM each dGTP, dATP, TTP, 30 μM [α-^{32}P]dCTP, oligo(dT) at 100 μg/ml, avian myeloblastosis virus reverse transcriptase (40 units), and 6 μg of size-fractionated poly(A)$^+$ RNA that had been first denatured with 1 mM methylmercury hydroxide and then treated with 55 mM 2-mercaptoethanol (49). After 1 hr at 42°C, the reaction was stopped by the addition of EDTA and then treated with 0.6 M NaOH overnight at room temperature and neutralized with 1.0 M HCl. The cDNA was separated from unincorporated nucleotides by chromatography on Sephadex G-50. Second strand synthesis was carried out by the sequential action of reverse transcriptase and the Klenow fragment of DNA polymerase (23). The conditions of incubation with reverse transcriptase were as described above for first strand synthesis, except that incubation was at 37°C for 90 min and [α-^{32}P]dCTP, and sodium pyrophosphate were omitted. Subsequent DNA synthesis with the Klenow fragment of *E. coli* DNA polymerase (20 units per μg of cDNA) was at 15°C for 3 hr. The DNA was recovered and treated with S1 nuclease (150 units per μg of cDNA) at 37°C for 60 min. After tailing with deoxycytidine (24), the cDNA was annealed to *Pst* I-digested, dG-tailed pBR322 (New England Nuclear) and cloned by transformation of *E. coli* RR1 (25).

Identification of β-Glucuronidase cDNA Clones. Colonies were transferred to duplicate sets of nitrocellulose filters (Schleicher & Schuell) and screened by differential hybridization to total and β-glucuronidase-enriched mRNA fractions labeled (10^7 cpm/μg) with [γ-^{32}P]ATP using T4 polynucleotide kinase (26). Hybridization was carried out at 65°C for 24 hr in 3× SSC/10× Denhardt's solution containing 0.02% NaDodSO$_4$, sonicated salmon sperm DNA (100 μg/ml), yeast tRNA (50 μg/ml), and 10^6 cpm of poly(A)$^+$ RNA per ml (1× SSC = 0.15 M NaCl/0.015 M sodium citrate; 1× Denhardt's solution = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone). Filters were washed several times at 65°C with 1× SSC containing 0.1% NaDodSO$_4$. Positive colonies were grown in liquid medium and plasmids were isolated by rapid alkaline extraction (27) or by urea/hydroxyapatite chromatography (28). The plasmid preparations were then tested for their ability to select translatable β-glucuronidase mRNA by hybridization selection (23), using 20–30 μg of rat preputial gland poly(A)$^+$ for filters containing 20 μg of immobilized plasmid DNA.

Construction of a Partial cDNA Library Using a Specific Oligonucleotide to Prime cDNA Synthesis. An oligonucleotide (30-mer) complementary to a sequence of β-glucuronidase mRNA (underlined in Fig. 3) derived from an incomplete cDNA clone was used as a primer. The rat preputial gland poly(A)$^+$ RNA (10 μg in 10 μl) was denatured with 2.5 mM methylmercury hydroxide for 10 min at room temperature and annealed to 10 pmol of the primer by incubation for 1 min at 90°C in a 20-μl reaction mixture containing 10 mM dithiothreitol, 200 mM KCl, 0.2 mM EDTA, followed by incubation for 10 min at 43°C. Cloned Moloney murine leukemia virus reverse transcriptase (1000 units) was then used for first strand cDNA synthesis in a 100-μl reaction mixture, according to the protocol of Bethesda Research Laboratories, except that 10–20 μCi of [α-^{32}P]dCTP was included (1 Ci = 37 GBq). Second strand synthesis and cloning into *E. coli* MC1016 cells with the plasmid vector pMG5 were carried out as described by Gubler and Hoffman (29). Recombinants were screened by colony hybridization with a cDNA restriction fragment from the insert in pβG5 as described (23), except that 0.3% NaDodSO$_4$ and salmon sperm DNA (250 μg/ml) were used in the hybridization solutions.

***In Vitro* Transcription–Translation of a Reconstructed β-Glucuronidase cDNA.** A cDNA containing the entire β-glucuronidase coding region inserted into the plasmid vector pSP64 was reconstructed in two steps: First, the 575-base-pair (bp) *Asu* II/*Pst* I fragment of pβG5 and the 700-bp *Pst* I/*Hpa* II fragment of pβG2 were ligated and introduced into the *Acc* I site of pUC9. The resulting plasmid contains a 1275-bp cDNA, which codes for the COOH-terminal 50% of β-glucuronidase. A 1200-bp *Hin*cII/*Eco*RI fragment isolated from this plasmid was ligated to an 1100-bp *Bam*HI/*Hin*cII fragment from pβG5′ and cloned into pSP64 digested with *Bam*HI and *Eco*RI. The *Bam*HI site in pβG5′ is located in the plasmid just upstream from the cDNA insert. Transcription of the reconstructed gene was carried out as described (30) with the following modifications: a final reaction volume of 10 μl contained ≈3 μg DNA, 20 mM Hepes (pH 7.4), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 1 mM each ribonucleotide, 1 mM $M^7G(5')ppp(5')G$, RNasin (2 units/μl), and SP6 polymerase (1 unit/μl). *In vitro*-transcribed RNA was translated in a wheat germ cell-free system (5).

RESULTS AND DISCUSSION

Isolation of Partial cDNA Clones for the Rat Preputial Gland β-Glucuronidase. Poly(A)$^+$ mRNA from rat preputial gland, enriched 20- to 30-fold in translatable β-glucuronidase mRNA by preparative electrophoresis in a denaturing agarose gel, was used to produce cDNA for cloning into pBR322 using the dG-dC tailing procedure (31). Of 500 clones obtained, 20 were selected for further study on the basis of their more intense hybridization to electrophoretically enriched mRNA than to unfractionated poly(A)$^+$ RNA. Three of these clones (pβG1, pβG2, and pβG5), with inserts between 500 and 1200 bp long, were shown to contain β-glucuronidase-specific sequences by hybridization-selection of translatable mRNA (Fig. 1).

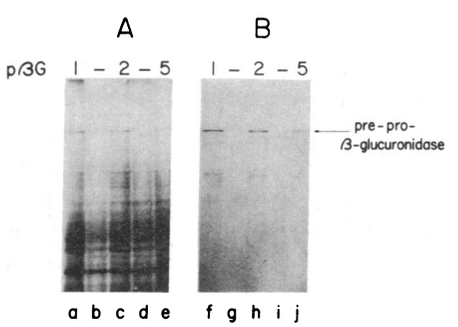

FIG. 1. Identification of β-glucuronidase cDNA clones by translation of hybridization-selected mRNA. The translation products of mRNA selected by hybridization to various clones identified by the preliminary colony hybridization procedure were analyzed by NaDodSO$_4$/polyacrylamide gel electrophoresis followed by autoradiography, either directly (*A*) or after immunoprecipitation with β-glucuronidase-specific antibodies (*B*). Only products of RNA selected by hybridization to clones pβG1, pβG2, and pβG5, analyzed in lanes a, c, and e, respectively, contained a polypeptide with the mobility of the primary translation product of β-glucuronidase mRNA, which was efficiently immunoprecipitated (lanes f, h, and j). The samples represented by lanes b, d, g, and i correspond to negative clones that serve as controls in this experiment.

Restriction endonuclease mapping showed that the cDNA segment in pβG1 was wholly contained with the longest clone (pβG2) whereas pβG5 appeared to only partially overlap with pβG2. DNA sequence analysis, by the strategy described in Fig. 2, indicated that a continuous cDNA sequence of ≈1.5 kilobases (kb) was contained in clones pβG2 and pβG5 and that this included an open reading frame corresponding to the carboxyl-terminal 334 amino acids of the encoded polypeptide, as well as 510 nucleotides from the 3′ untranslated region of mRNA (Figs. 2 and 3). Since the molecular mass of the β-glucuronidase subunit is ≈75 kDa (4), the open reading frame within the two overlapping clones must represent ≈50% of the β-glucuronidase mRNA coding region. Indeed, the size of the preputial gland mRNA encoding β-glucuronidase was shown to be 2.6 kb by RNA blot analysis, using as a probe a ^{32}P-labeled plasmid containing the recombined pβG2 and pβG5 inserts (Fig. 4).

To obtain cDNA clones encompassing the remainder of the mRNA sequence, a second cDNA library was constructed from rat preputial gland total poly(A)$^+$ RNA using as a specific primer for cDNA synthesis an oligonucleotide complementary to the 5′ end of the previously determined mRNA sequence. Nineteen clones containing inserts varying in size from 450 to 1350 bp were identified by hybridization with a ^{32}P-labeled *Hin*cII/*Cla* I fragment of pβG5 (Fig. 2). The clone with the largest insert (pβG5′) was sequenced (as outlined in Fig. 2) and found to contain a 1350-bp cDNA segment with an open reading frame that overlapped at its 3′ end with the 5′ portion of pβG5.

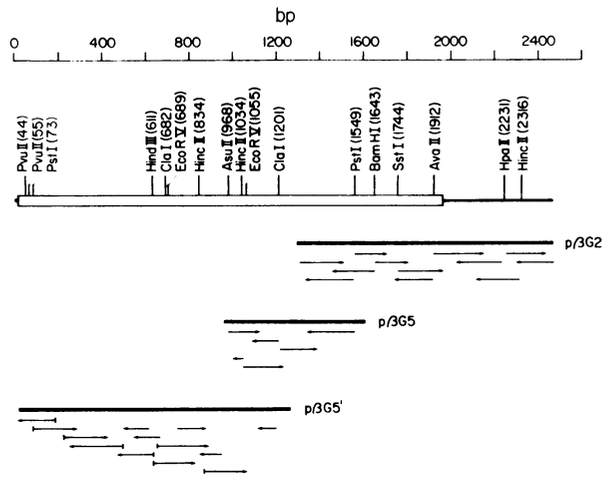

FIG. 2. Restriction map of β-glucuronidase cDNA clones and sequencing strategy. Restriction maps of the cDNA clones pβG2, pβG5, and pβG5′ were obtained by conventional procedures (23) and a composite map is shown. All restriction sites were verified by subsequent DNA sequencing. The coding region is represented by the open bar and the 3′ and 5′ untranslated regions are represented by the heavy line. Arrows indicate the direction and extent of each sequence determination. Those arrows with vertical lines at their tails indicate regions sequenced by a modification of the dideoxy-chain determination method (32) using as primers the M13 universal primer or synthetic oligonucleotides chosen on the basis of previously determined sequences. Arrows without vertical lines at their tails indicate regions sequenced by the Maxam–Gilbert method (33).

```
                -22                                                -11
          MET SER PRO ARG ARG SER VAL CYS TRP PHE VAL LEU GLY GLN LEU LEU CYS SER
CCC CGC ATG TCC CCA AGA CGG AGT GTG TGT TGG TTC GTA CTG GGC CAG CTG CTG TGC AGC

                    1                                              10
CYS ALA VAL ALA LEU GLN GLY GLY MET LEU PHE PRO LYS GLU THR PRO SER ARG GLU LEU
TGC GCC GTG GCG CTG CAG GGC GGG ATG CTG TTC CCG AAG GAG ACC CCT TCG CGG GAA CTC

            20                                                 30
LYS VAL LEU ASP GLY LEU TRP SER PHE ARG ALA ASP TYR SER ASN ASN ARG LEU GLN GLY
AAG GTG CTG GAC GGC CTG TGG AGC TTC CGT GCC GAT TAC TCG AAC AAT CGG TTG CAG GGC

            40                                                 50
PHE GLU LYS GLN TRP TYR ARG GLN PRO LEU ARG GLU SER GLY PRO THR LEU ASP MET PRO
TTC GAG AAG CAG TGG TAC CGG CAG CCG CTA CGG GAG TCG GGC CCA ACC TTG GAC ATG CCG

            60                                                 70
VAL PRO SER SER PHE ASN ASP ILE THR GLN GLU ALA GLU LEU ARG ASN PHE ILE GLY TRP
GTC CCT TCC AGC TTC AAT GAC ATC ACC CAA GAA GCA GAG CTT CGG AAC TTC ATT GGC TGG

            80                                                 90
VAL TRP TYR GLU ARG GLU ALA VAL LEU PRO GLN ARG TRP THR GLN ASP THR ASP ARG ARG
GTG TGG TAT GAA CGG GAA GCA GTG CTT CCA CAG CGA TGG ACC CAG GAC ACC GAC AGG AGA

           100                                                110
VAL VAL LEU ARG ILE ASN SER ALA HIS TYR TYR ALA VAL VAL TRP VAL ASN GLY ILE HIS
GTG GTG TTG AGA ATC AAC AGC GCC CAT TAC TAT GCA GTT GTG TGG GTG AAT GGG ATT CAT

           120                                                130
VAL VAL GLU HIS GLU GLY GLY HIS LEU PRO PHE GLU ALA ASP ILE THR LYS LEU VAL GLN
GTG GTG GAA CAT GAG GGA GGT CAC CTC CCC TTT GAG GCT GAC ATC ACC AAG CTG GTC CAG

           140                                                150
SER GLY PRO LEU THR THR PHE ARG VAL THR ILE ALA ILE ◆ASN ASN THR LEU THR PRO TYR
AGT GGG CCC CTG ACC ACC TTC CGG GTC ACC ATC GCC ATC AAC AAC ACA CTG ACC CCT TAT

           160                                                170
THR LEU PRO PRO GLY THR ILE VAL TYR LYS THR ASP PRO SER MET TYR PRO LYS GLY TYR
ACC CTT CCA CCG GGG ACC ATT GTC TAC AAG ACT GAT CCT TCC ATG TAT CCC AAG GGT TAC

           180                                                190
PHE VAL GLN ASP ILE SER PHE ASP PHE PHE ASN TYR ALA GLY LEU HIS ARG SER VAL VAL
TTC GTC CAG GAC ATA AGC TTC GAC TTC TTC AAC TAT GCG GGG CTG CAC CGG TCT GTG GTC

           200                                                210
LEU TYR THR THR PRO THR THR TYR ILE ASP ASP ILE THR VAL THR THR ASP VAL ASP ARG
CTG TAC ACC ACC CCT ACC ACC TAT ATC GAT GAT ATC ACT GTG ACC ACT GAC GTG GAC CGG

           220                                                230
ASP VAL GLY LEU VAL ASN TYR TRP ILE SER VAL GLN GLY SER ASP HIS PHE GLN LEU GLU
GAC GTT GGG CTG GTG AAC TAC TGG ATT TCT GTC CAG GGC AGT GAC CAT TTC CAG CTA GAA

           240                                                250
VAL ARG LEU LEU ASP GLU ASP GLY LYS ILE VAL ALA ARG GLY THR GLY ASN GLU GLY GLN
GTG CGT CTT CTG GAT GAG GAT GGC AAA ATT GTG GCC CGT GGA ACA GGG AAT GAG GGT CAA

           260                                                270
LEU LYS VAL PRO ARG ALA HIS LEU TRP TRP PRO TYR LEU MET HIS GLU HIS PRO ALA TYR
CTT AAG GTG CCG CGT GCC CAC CTC TGG TGG CCT TAC CTG ATG CAT GAG CAT CCA GCC TAC

           280                                                290
LEU TYR SER LEU GLU VAL THR MET THR THR PRO GLU SER VAL SER ASP PHE TYR THR LEU
CTG TAC TCC TTG GAG GTG ACG ATG ACA ACA CCT GAG TCT GTG TCT GAC TTC TAT ACC CTC

           300                                                310
PRO VAL GLY ILE ARG THR VAL ALA VAL THR LYS SER LYS PHE LEU ILE ASN GLY LYS PRO
CCT GTC GGG ATT CGA ACA GTG GCT GTC ACA AAG AGC AAG TTC CTC ATA AAT GGG AAG CCT

           320                                                330
PHE TYR PHE GLN GLY VAL ASN LYS HIS GLU ASP SER ASP ILE ARG GLY ARG GLY PHE ASP
TTC TAC TTC CAA GGC GTC AAC AAG CAT GAG GAT TCA GAT ATC CGA GGG AGA GGC TTC GAC
```

```
           340                                                350
TRP PRO LEU LEU ILE LYS ASP PHE ASN LEU LEU ARG TRP LEU GLY ALA ASN SER PHE ARG
TGG CCT CTG CTG ATA AAG GAT TTC AAC CTC CTC CGT TGG CTC GGG GCA AAT TCC TTT CGT

           360                                                370
THR SER HIS TYR PRO TYR SER GLU GLU VAL LEU GLN LEU CYS ASP ARG TYR GLY ILE VAL
ACC AGC CAC TAT CCC TAC TCG GAG GAG GTA CTT CAG CTC TGT GAC CGA TAT GGA ATT GTG

           380                                                390
VAL ILE ASP GLU CYS PRO GLY VAL GLY ILE VAL LEU PRO GLN SER PHE GLY ◆ASN VAL SER
GTC ATC GAT GAG TGT CCC GGT GTG GGC ATC GTG CTG CCC CAG AGT TTT GGC AAC GTG TCT

           400                                                410
LEU ARG HIS HIS LEU GLU VAL MET ASP GLU LEU VAL ARG ARG ASP LYS ASN HIS PRO ALA
CTT CGG CAC CAC CTA GAG GTG ATG GAC GAG CTG GTG CGC AGG GAC AAA AAT CAC CCT GCG

           420                                                430
VAL VAL MET TRP SER VAL ALA ASN GLU PRO VAL SER SER LEU LYS PRO ALA GLY TYR TYR
GTC GTG ATG TGG TCT GTG GCC AAT GAG CCT GTC TCT TCT CTG AAA CCT GCC GGA TAT TAC

           440                                                450
PHE LYS THR LEU ILE ALA HIS THR LYS ALA LEU ASP PRO THR ARG PRO VAL THR PHE VAL
TTC AAG ACG CTG ATC GCC CAC ACC AAA GCC CTG GAC CCC ACC CGT CCT GTG ACC TTT GTG

           460                                                470
SER ASN THR ARG TYR ASP ALA ASP MET GLY ALA PRO TYR VAL ASP VAL ILE CYS VAL ASN
AGC AAT ACC AGA TAT GAC GCA GAC ATG GGG GCC CCG TAC GTG GAC GTG ATT TGT GTG AAC

           480                                                490
SER TYR LEU SER TRP TYR HIS ASP TYR GLY HIS LEU GLU VAL ILE GLN LEU GLN LEU THR
AGT TAC TTA TCC TGG TAT CAT GAC TAC GGG CAT CTG GAG GTG ATT CAG CTG CAG CTG ACT

           500                                                510
SER GLN PHE GLU ASN TRP TYR LYS MET TYR GLN LYS PRO ILE ILE GLN SER GLU TYR GLY
AGC CAG TTT GAG AAC TGG TAT AAG ATG TAC CAG AAG CCA ATT ATC CAG AGC GAG TAT GGA

           520                                                530
ALA ASP ALA VAL SER GLY LEU HIS GLU ASP PRO PRO ARG MET PHE SER GLU GLU TYR GLN
GCA GAC GCC GTC TCG GGG CTT CAT GAG GAT CCA CCT CGC ATG TTC AGT GAG GAG TAC CAG

           540                                                550
THR ALA LEU LEU GLU ASN TYR HIS LEU ILE LEU ASP GLU LYS ARG LYS GLU TYR VAL ILE
ACA GCT CTC CTG GAG AAT TAT CAT TTG ATC CTG GAT GAG AAA CGA AAA GAA TAT GTC ATC

           560                                                570
GLY GLU LEU ILE TRP ASN PHE ALA ASP PHE MET THR ◆ASN GLN SER PRO LEU ARG VAL THR
GGA GAG CTC ATC TGG AAT TTT GCT GAC TTC ATG ACG AAC CAG TCA CCA CTG AGA GTA ACA

           580                                                590
GLY ASN LYS LYS GLY ILE PHE THR ARG GLN ARG ASN PRO LYS MET ALA ALA PHE ILE LEU
GGA AAC AAG AAG GGG ATC TTC ACT CGA CAG AGA AAC CCC AAG ATG GCA GCC TTC ATT TTG

           600                                                610
ARG GLU ARG TYR TRP ARG ILE ALA ◆ASN GLU THR ARG GLY TYR GLY SER VAL PRO ARG THR
CGA GAG AGA TAC TGG AGG ATT GCC AAC GAA ACC AGA GGT TAC GGT TCA GTG CCG AGG ACC

           620
GLN CYS MET GLY SER ARG PRO PHE THR PHE ---
CAG TGT ATG GGA AGC AGA CCG TTC ACA TTC TAA AGCTACGACTACCTCACCGCAGACAGGGCCTGATGA

AGTGAAAGGTATTCTCTCAGCAGAAGAGTTTTTCCTCTGGAGTTTGTGAACCAACACAAGGAGGACTTTAGAAGTGACAT

GGAAGTTTTCCTTGTGTGGGGATAATGACTTGCAGTGCCTTAGGCTCAGTAATGTGTCCTGCTGAGAAGTGTCCCAGGTC

ATTTCTGTCAGGTTTTTTTTGGCCTTGGCTTTGTGTACTCTTGAAAACCTACTGTGTGAATACTCACACTAAGAGCTCCT

GGTTACCGGCTAAGGGACGGCTCACAATTTAAGAATACTGACTGTTCTACCAGGGGTTCCAGATTCGATTCCTAGCACCC

ACGTGGTCGTTAACTATGAGCTGTTGCCCCATTTCTCAGGTGATGTACGAACATACATGTGGGGGGAGGGGAATAGTCAT

ACGTGTAAAATGAAAATAAATCTAAATAAAACTGTTTTCTTTCATTATAAAAAAAAAAAAAAAAAACCAACTAATAGAAA

CGT
```

FIG. 3. Nucleotide sequence of β-glucuronidase cDNA and amino acid sequence of the encoded polypeptide. The arrow indicates the cleavage site for the signal peptidase and residue 1 represents the first amino acid of the mature β-glucuronidase. The diamonds indicate asparagine residues that could serve as acceptors for oligosaccharide chains. The sequence complementary to the primer used to generate clone pβG5′ is underlined. Two polyadenylylation signals (overlined) are present near the 3′ end of the cDNA but the 18-residue-long poly(A) segment present in the sequence is not at the extreme 3′ end. It is therefore likely that the cDNA does not extend to the end of the mRNA.

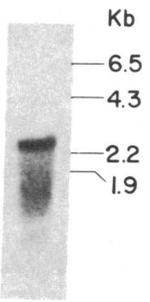

FIG. 4. Hybridization of rat preputial gland mRNA to cloned β-glucuronidase cDNA. Poly(A)$^+$ RNA (20 μg) from rat preputial gland was denatured with glyoxal and electrophoresed in a 1% agarose gel (23). The RNA was transferred to a nylon filter (Biodyne A Transfer Membrane) and hybridized (24) to a ^{32}P-labeled plasmid containing the inserts in pβG2 and pβG5 recombined at their common *Pst* I site (see Fig. 2). The positions of DNA molecular size markers are indicated.

The composite nucleotide sequence obtained from the overlapping clones by the strategy described in Fig. 2 and the predicted amino acid sequence for the entire β-glucuronidase polypeptide are shown in Fig. 3. An open reading frame of 1944 nucleotides begins with an ATG codon 2 nucleotides downstream from the 5′ end of pβG5′. This reading frame codes for a protein of 648 amino acids with a calculated molecular mass of 74.8 kDa, which is slightly higher than that deduced from the electrophoretic mobility of the unglycosylated primary translation product of rat preputial gland mRNA (5). The sequence of the first 10 amino-terminal residues of purified rat preputial gland β-glucuronidase was determined by automated Edman degradation (data not shown) and found to correspond to residues 23–32 of the polypeptide encoded by the cDNA. It can be concluded, therefore, that β-glucuronidase contains a transient amino-terminal signal sequence of 22 amino acid residues. It should also be noted that, like other lysosomal enzymes, β-glucuronidase undergoes a posttranslational proteolytic cleavage near its carboxyl-terminal end (34), which reduces its size by ≈3 kDa. The site of this cleavage has not yet been determined. The sequences of several cysteine-containing peptide fragments derived from rat preputial gland β-glucuronidase have been reported (35). Two of these peptide sequences are found within the polypeptide sequence derived by us at residues 370–377 and 470–478. The last 10 residues of a third reported peptide of 11 residues are also present at residues 376–385, overlapping with one of the other two perfectly matching segments. The other three cysteine-containing peptides previously reported (35) are not found within the β-glucuronidase sequence deduced from the cDNA clone. The sequence of the rat preputial gland cDNA was compared to the recently reported sequences of 188-bp restriction fragments of mouse and human β-glucuronidase cDNAs (20). Residues 494–559 of the rat cDNA showed 85% and 79% homology to the mouse and human cDNA fragments, respectively.

The sequence predicted for the rat preputial gland enzyme (Fig. 3) contains four possible N-glycosylation sites. Three N-linked oligosaccharide chains have been identified in the β-glucuronidase produced by a murine macrophage cell line (8), but the number of oligosaccharide chains present in the rat preputial gland enzyme has not been established.

A search for amino acid sequence homology between β-glucuronidase and other proteins in the National Biomedical Research Foundation data base revealed that a region of β-glucuronidase extending from residues 101 to 477 has 23% identity with a portion of the *E. coli* β-galactosidase extending from residues 123 to 486 (Fig. 5). Within this region, there are three short segments of substantially higher homology in

```
 99' VWYEREAVLPQRWTQDTDRRVVLRINSAHYYAVVWVNGIHVVEHEGGHLPFEADITKLVQSGPLTT
121" GCYSLTFNVDESWLQEGQTRIIFDGVNSAFH--LWCNGRWVGYGQDSRLPSEFDLSAFLRAGE-NR

165' FRVTIAINNTLTPYTLPPGTIVYKTDPSMYPKGYFVQDISFDFFNYAGLHRSVVLYTTPTTYIDDI
184" LAVMV-LRWSDGSY-L-EDQDMWRM-SGIFRDVSLLHKPTTQISDFHVATRFNDDF-SRAVLEAEV

231' TVTTDVDRDVGL-VNYW--ISVQGSDHFQLEVRLLDEDGKIVARGTGNEGQLKVPRAHLWWPYLMH
245" QMCGELRDYLRVTVSLWQGETQVASGTAPFGGEIIDERGGYADRVT---LRLNVENPKLWSA----

294' EHPAYLYSLEVTMTTPE-SVSDFYTLPVGIRTVAVTKSKFLINGKPFYFQGVNKHEDSDIRGRGFD
304" EIPN-LYRAVVELHTADGTLIEAEACDVGFREVRIENGLLLLNGKPLLIRGVNRHEHHPLHGQVMD

359' WPLLIKDFNLLRWLGANSFRTSHYPYSEEVLQLCDRYGIVVIDE--CPGVGIVLPQSFGNVSLR-H
369" EQTMVQDILLMKQNNFNAVRCSHYPNHPLWYTLCDRYGLYVVDEANIETHGMVPMNRLTDDPRWLP

422' HL-EVMDELVRRDKNHPAVVMWSVANEPVSSLKPAGYYFKTLIAHTKALDPTRPVTF
435" AMSERVTRMVQRDRNHPSVIIWSLGNESGHGANHDALY-----RWIKSVDPSRPVQY
```

FIG. 5. Homology between rat preputial gland β-glucuronidase and *E. coli* β-galactosidase. The National Biomedical Research Foundation data base (updated to May 1986) was searched for homologies using the FASTP program (36). All sequences in the data base gave an initial similarity score of 56, except *E. coli* β-galactosidase, which gave an initial similarity score of 165. The optimized alignment for the two sequences (36) shown gave an optimized score of 266. These similarity scores are higher than those found when the bovine cAMP-dependent protein kinase was compared to the sequences of all available oncogene protein kinases (36). The initial score for the comparison of β-glucuronidase to β-galactosidase was higher than those scores observed when the bovine trypsinogen sequence was compared to those of many members of the serine protease superfamily. A colon between aligned residues in the two sequences indicates identity, whereas a single dot represents a conservative replacement as defined by the PAM250 replaceability matrix constructed by Dayhoff *et al.* (37). Boxed sequences represent regions of substantial homology (see text). Amino acids are identified by the single-letter code.

which 10 of 16, 9 of 12, and 11 of 18 residues match exactly (Fig. 5). These homologies may reflect the common structural features necessary for the activity of both glycosidases. Chemical inactivation studies of β-galactosidase have indicated the involvement of three specific residues (methionine-502, ref. 38; tyrosine-253, ref. 39, glutamic acid-461, ref. 40) in the enzyme active site. Only one of these residues is conserved in β-glucuronidase (glutamic acid-448), where it is part of a short segment of exceptionally high homology to β-glucuronidase (11 identical residues of 18, with the other 7 representing conservative replacements). The presence of a large region of homology suggests an evolutionary relationship between the bacterial and eukaryotic enzymes.

No significant homology was found between β-glucuronidase and the recently reported sequences of other lysosomal enzymes, including rat cathepsins B and H (41, 42), porcine (43, 44) and human (45) cathepsin D, human fucosidase (46), glucosecerebrosidase (47), and β-hexosaminidase (48). The lack of significant homology between the different lysosomal enzymes suggests that the recognition signal for the phosphorylation of mannose residues is not determined solely by the primary amino acid sequence.

***In Vitro* Expression of β-Glucuronidase cDNA.** A cDNA containing the complete coding region for β-glucuronidase was reconstructed from pβG2, pβG5, and pβG5′, as described in the legend to Fig. 6 and cloned into the plasmid vector pSP64. Transcription of the resulting recombinant plasmid with SP6 RNA polymerase yielded synthetic mRNA that was used to program a wheat germ cell-free translation system. The major translation product obtained had an electrophoretic mobility corresponding to that of the primary translation product (70 kDa) of preputial gland β-glucuronidase mRNA (5) and was immunoprecipitated by anti-β-glucuronidase antiserum (Fig. 4, lane a). When translation was performed in the presence of microsomal membranes,

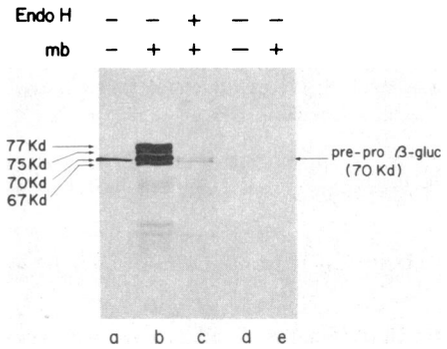

FIG. 6. Characterization of the product encoded by a reconstructed β-glucuronidase cDNA: Cotranslational processing and transfer into the microsomal lumen. Approximately 0.3 μg of RNA obtained by *in vitro* transcription of the reconstructed β-glucuronidase cDNA was translated in the absence of dog pancreas microsomes (mb), and the products were immunoprecipitated with anti-β-glucuronidase antiserum (lane a) or nonimmune serum (lane d). Approximately 1.2 μg of RNA was translated in the presence of microsomes (mb) and one-third of this translation mixture was used directly for immunoprecipitation with anti-β-glucuronidase antiserum (lane b). One-third of the sample was incubated with endoglycosidase H before immunoprecipitation (lane c) and the remaining was incubated with nonimmune serum (lane e). The molecular mass of the different polypeptides is indicated at the left.

several products with mobilities corresponding to 67, 70, 75, and 77 kDa were obtained (Fig. 6, lane b). The additional bands obtained in the presence of membranes must represent β-glucuronidase polypeptides that are cotranslationally inserted into the membranes, and undergo cleavage of the insertion signal with the addition of a variable number of oligosaccharide chains. The 67-kDa polypeptide probably represents β-glucuronidase that was incorporated into the microsomes, and hence had its signal cleaved, but was not glycosylated. The presence of oligosaccharide chains in the 75- and 77-kDa polypeptides was confirmed by the finding that treatment with endoglycosidase H eliminated the corresponding bands (Fig. 6, lane c). It is clear from these experiments that the reconstructed cDNA clone codes for the entire β-glucuronidase polypeptide.

The availability of a cDNA clone containing the entire β-glucuronidase coding sequence has enabled us to construct mutant derivatives and to obtain their expression in transfected cells. These studies should lead to the identification of the segments that are important for recognition of the polypeptide by the Golgi phosphotransferase that phosphorylates mannose residues within the core oligosaccharides and creates the sorting signals that leads to its sequestration within lysosomes. Current transfection experiments (R.A., unpublished observations) using the reconstructed cDNA inserted into a simian virus 40 based eukaryotic expression vector indicate that the encoded polypeptide is phosphorylated and incorporated into lysosomes.

We thank Mort Schneider for his assistance with the DNA sequencing; H. Pleskin, B. Zeitlow, and J. Culkin for photography; and M. Cort and B. Rosen for typing the manuscript. This work was supported by Grants GM 20277 and AG 01461 from the National Institutes of Health. R.A. is a fellow of the Jane Coffin Childs Memorial Fund.

1. McKusick, V. A. & Neufeld, G. F. (1983) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. D., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), pp. 751–777.
2. Tomino, S. & Paigen, K. (1975) *J. Biol. Chem.* **250,** 1146–1148.
3. Levy, G. A., McAllan, A. & Marsh, C. A. (1958) *Biochem. J.* **69,** 22–27.
4. Himeno, J., Ohara, H., Arakawa, Y. & Kato, N. (1975) *J. Biochem. (Tokyo)* **77,** 427–430.
5. Rosenfeld, M. G., Kreibich, G., Popov, D., Kato, K. & Sabatini, D. D. (1982) *J. Cell Biol.* **93,** 135–143.
6. Erickson, A. H. & Blobel, G. (1979) *J. Biol. Chem.* **254,** 11771–11774.
7. Tabas, I. & Kornfeld, S. (1980) *J. Biol. Chem.* **255,** 6633–6639.
8. Goldberg, D. E. & Kornfeld, S. (1981) *J. Biol. Chem.* **256,** 13060–13067.
9. Weismann, U. N., Lightbody, J., Vassella, F. & Herschkowitz, N. N. (1971) *N. Engl. J. Med.* **284,** 109–110.
10. Bach, G., Bargal, R. & Contz, M. (1979) *Biochem. Biophys. Res. Commun.* **91,** 976–981.
11. Hasilik, A. & Neufeld, E. F. (1980) *J. Biol. Chem.* **255,** 4937–4950.
12. Hasilik, A., Voss, B. & Von Figura, K. (1981) *Exp. Cell Res.* **133,** 23–30.
13. Reitman, M. L., Varki, A. & Kornfeld, S. (1981) *J. Clin. Invest.* **67,** 1574–1579.
14. Varki, A., Reitman, M. L., Vannier, A., Kornfeld, S., Grubb, J. H. & Sly, W. (1982) *Am. J. Hum. Genet.* **34,** 717–729.
15. Creek, K. E. & Sly, W. S. (1984) in *Lysosomes in Biology and Pathology*, eds. Dingle, J. T., Dean, R. T. & Sly, W. (Elsevier, New York), pp. 45–60.
16. Hieber, V. C. (1982) *Biochem. Biophys. Res. Commun.* **104,** 1271–1278.
17. 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.
18. Palmer, R., Gallagher, P. M., Boyko, W. L. & Genschow, R. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7596–7600.
19. Catterall, J. F. & Leary, S. L. (1983) *Biochemistry* **22,** 6049–6053.
20. 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.
21. Chirgwin, J. M., Przybla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **24,** 5294–5299.
22. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
24. Roychoudhury, R., Jay, E. & Wu, R. (1976) *Nucleic Acids Res.* **3,** 863–877.
25. Cohen, S. N., Chang, A. C. Y. & Hsu, L. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 2110–2114.
26. Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M. & Darnell, J. E., Jr. (1981) *Cell* **23,** 731–739.
27. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
28. Colman, A., Byers, M. J., Primrose, J. B. & Lyons, A. (1978) *Eur. J. Biochem.* **91,** 303–310.
29. Gubler, U. & Hoffman, B. (1983) *Gene* **25,** 263–269.
30. Melton, D. A., Kreig, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* **12,** 7035–7056.
31. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naker, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3727–3731.
32. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
33. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
34. Erickson, A. H. & Blobel, G. (1983) *Biochemistry* **22,** 5201–5205.
35. Leighton, P. H., Fisher, W. K., Moon, K. E. & Thompson, E. O. P. (1980) *Aust. J. Biol. Sci.* **33,** 513–520.
36. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
37. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345–352.
38. Fowler, A. V., Zabin, I., Sinnott, M. L. & Smith, P. J. (1978) *J. Biol. Chem.* **253,** 5283–5285.
39. Huber, R. E., Fowler, A. V. & Zabin, I. (1982) *Biochemistry* **21,** 5052–5055.
40. Herrchen, M. & Legler, G. (1984) *Eur. J. Biochem.* **138,** 527–531.
41. Takio, K., Towatori, T., Katunuma, N., Teller, O. C. & Titani, K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3666–3670.
42. Segundo, B. S., Chan, S. J. & Steiner, D. F. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2320–2324.
43. Takahashi, T. & Tang, J. (1983) *J. Biol. Chem.* **258,** 6435–6438.
44. Shewale, J. G. & Tang, J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3703–3707.
45. Faust, P. L., Kornfeld, S. & Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4910–4914.
46. Fukushima, H., DeWit, J. R. & O'Brien, J. S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1262–1265.
47. Tsuji, S., Choudary, P. V., Martin, B. M., Winfield, S., Barranger, J. A. & Ginns, E. I. (1986) *J. Biol. Chem.* **261,** 50–53.
48. Myerowitz, R., Piekarz, R., Neufeld, E. F., Shows, T. B. & Kunihiko, S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7830–7834.
49. Bailey, J. M. & Davidson, N. (1976) *Anal. Biochem.* **70,** 75–85.