Human α-N-Acetylglucosaminidase-Molecular Cloning, Nucleotide Sequence, and Expression of a Full-length cDNA

HOMOLOGY WITH HUMAN α-GALACTOSIDASE A SUGGESTS EVOLUTION FROM A COMMON ANCESTRAL GENE

(Received for publication, April 20, 1990)

Anne M. Wangt, David F. Bishop, and Robert J. Desnick§

From the Division of Medical and Molecular Genetics, Mount Sinai School of Medicine, New York, New York 10029

Human α-N-acetylglucosaminidase (α-GalNAc, E.C. 3.2.1.49), the lysosomal glycohydrolase that cleaves α-N-acetylglucosaminyl moieties from glycoconjugates, is encoded by a gene localized to chromosome 22q13-qter. The deficient activity of α-GalNAc is the enzymatic defect in Schindler disease, an inherited neuroaxonal dystrophy. To isolate a full-length cDNA, the enzyme from human lung was purified to homogeneity, 129 non-overlapping amino acids were determined by microsequencing the N terminus and seven tryptic peptides, and synthetic oligonucleotide mixtures were used to screen a human fibroblast cDNA library. A full-length cDNA, pAGB-3, isolated from a placental λgt11 cDNA library, had a 1585-base pair (bp) insert with an open reading frame which predicted an amino acid sequence that was colinear with all 129 microsequenced residues of the purified enzyme. The pAGB-3 insert had a 3143-bp 5'-untranslated region, a 1236-bp open reading frame encoding 411 amino acids, a 514-bp 3'-untranslated region, and a 64-bp poly(A) tract. A signal peptide sequence of 17 amino acids as well as six N-glycosylation sites were predicted. The pAGB-3 cDNA was subcloned into the pBluescript(B) mammalian expression vector and human α-GalNAc activity was transiently expressed in COS-1 cells, demonstrating the functional integrity of the full-length cDNA. Northern hybridization analysis of mRNA revealed two transcripts of about 3.6 and 2.2 kilobases (kb), and primer extension studies indicated a cap site at nucleotide -347 for the 2.2-kb transcript. The 3.6-kb cDNA (pAGB-35) was isolated; the 3598-bp pAGB-35 insert was identical to that of the 2.2-kb insert but had additional 5'- and 3'-untranslated sequences including a second downstream polyadenylation signal at nucleotide 3100–3105. Isolation of a genomic clone, gAGB-1, and sequencing the 2046-bp region including pAGB-3 revealed a 1754-bp intron between codons 319 and 320, which also was the site of a 70-bp insertion and a 45-bp deletion in other cDNA clones. Notably, the α-GalNAc cDNA had remarkable amino acid homology with the human α-galactosidase A (α-Gal A) cDNA suggesting the evolutionary relatedness of these genes. The α-GalNAc cDNA had 46.9–64.7% amino acid identity in sequences (codons 1–319) corresponding to α-Gal A exons 1 through 6, while the comparable exon 7 sequence (pAGB-3 codons 320–411) had only 15.8% homology with numerous gaps. These findings implicate the genomic region at and surrounding codon 319 as a potential site for the abnormal processing of α-GalNAc transcripts as well as for a recombinational event in the evolution and divergence of α-Gal A and α-GalNAc.

The availability of the full-length cDNA for human α-GalNAc will permit studies of the genomic organization and evolution of this lysosomal gene, as well as the characterization of the molecular lesions causing Schindler disease.

In the early 1970s, several investigators demonstrated the existence of two α-galactosidase isozymes, designated A and B, which hydrolyzed the α-galactosidic linkages in 4-MU-α-D-galactopyranosides (1–7). In tissues, about 80–90% of total α-galactosidase (α-Gal) activity was due to a thermolabile, myoinositol-inhibitable α-Gal A isozyme, while a relatively thermostable, α-Gal B, accounted for the remainder. The two “isozymes” were separable by electrophoresis, isoelectric focusing, and ion-exchange chromatography. After neuraminidase treatment, the electrophoretic migrations and pI values of α-Gal A and B were very similar (1), initially suggesting that the two enzymes were the differentially glycosylated products of the same gene. The finding that the purified glycoprotein enzymes had similar physical properties including subunit molecular mass (∼40 kDa), homodimeric structures, and amino acid compositions also indicated their structural relatedness (8–14). However, the subsequent demonstration that polyclonal antibodies against α-Gal A or B did not cross-react with the other enzyme (8, 11), that only α-Gal A activity was deficient in hemizygotes with Fabry disease (1–8), and that the genes for α-Gal A and B mapped to different chromosomes (7, 15), clearly demonstrated that these enzymes were genetically distinct. Thus, it was not surprising when α-Gal B was shown in 1977 to be an α-N-acetylgalactosaminidase (α-GalNAc), a homodimeric

1 The abbreviations used are: 4-MU- or pNP-α-GalNAc, 4-methylumbelliferyl- or p-nitrophenyl-α-N-acetylgalactosaminide; α-Gal, α-galactosidase; α-GalNAc, α-N-acetylgalactosaminidase; bp, base pair; 4-MU-α-Gal, 4-methylumbelliferyl-α-D-galactoside; nt, nucleotide; PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; HPLC, high performance liquid chromatography; kb, kilobase(s).
Full-length cDNA Encoding Human α-N-Acetylgalactosaminidase

EXPERIMENTAL PROCEDURES

Affinity Purification, Microsequencing, and Antibody Production—Human lung α-GalNAc was purified to homogeneity, polyclonal rabbit anti-human α-GalNAc antibodies were produced and purified, and cell supernatants were immunoblotted as described previously (18, 21, 22). For the isolation of tryptic peptides, the concentrated post-hydroxylapatite fraction was subjected to preparative 10% NaDodSO4/PAGE and the 48- and 117-kDa species were determined by automated gas-phase microsequencing and HPLC identification of phenylthiohydantoin amino acid derivatives (25).

Construction of Synthetic Oligonucleotide Probes—Mixed and unique oligonucleotides were synthesized on an Applied Biosystems model 380B oligonucleotide synthesizer. Four oligonucleotide mixtures were constructed to regions of minimal codon redundancy for the following nucleic acid fragments: N terminal (5'-CA(A)AGCNCCCNATGGG-3'), peptide T-106A (AA(T)AT(TCA)GA(TC)GTG(TTCTGATTT(CAAGG-3'), and peptide T-133, (5'-TGGCCNGNTATCGA(AGGGG-3'). Oligonucleotide probes for library screening were 5'-end-labeled with [γ-32P] ATP using T4 polynucleotide kinase (26). Unique sequence oligonucleotides (17-mers) were synthesized and used as primers in sequencing reactions. To determine if the same site, two unique, overlapping 30-mer probes were synthesized for primer extension, 5'-TGCGGACCTCCAGACAGGATGTTGTA-3' and 5'-CTCGAGGAGTTGTTAGGAGCTCACTGG-3'. To detect alternatively spliced transcripts, PCR sense and antisense primers for the exonic region flanking the putative 70-bp insertion had the sequences 5'-AGTGCAATTCTCTGACACCGTGGG-3' and 5'-AGTCATGGCATGACCAGGCTG-3', respectively. The four PCR primer sequences for the construction of the α-Gal A and α-Galα hybrid cDNA were α-Gal A sense, 5'-TGGGAGTAGATGCTCGTCTAAAA-3'; α-Gal A antisense, 5'-GATGAGAGTTTCTCTCTTCTAGTTGGT-3'; α-Galα sense, 5'-TACCAGTTCGAGCGAGAACACCATGCT-3'; and α-Galα antisense, 5'-AAGGAGTCCAGATCTCTTACT-3'.

Isolation and Characterization of cDNA and Genomic Clones—The human fibroblast cDNA library, kindly provided by Dr. Hiroto Okayama (National Institutes of Health), was screened with the radiolabeled 26-mer oligonucleotide mixture corresponding to tryptic peptide T-106A by colony hybridization (24). Plasmid DNA isolation and Southern analyses of positive clones were performed as previously described (26). For isolation of a full-length cDNA, a 0.9-kb BamHI fragment corresponding to the 5' portion of the pAGB-1 insert was then isolated, nick-translated, and used to screen recombinants from a λgt11 human placental library (Clontech Laboratories, Palo Alto, CA) by plaque hybridization (26). A λgt10 human retinal cDNA library, kindly provided by Dr. Jeremy Nathans (Johns Hopkins University), was screened with the radiolabeled pAGB-3 cDNA insert using the conditions described above. To isolate genomic clones containing the entire α-GalNAc sequence, 1 × 106 recombinants from a human genomic cosmid library were screened with the radiolabeled pAGB-5 cDNA insert using the conditions described above for cDNA library screening. The genomic library was prepared from size-selected human lymphoblast DNA and kindly provided by Dr. Henrik Vissing (Mount Sinai School of Medicine). DNA Sequencing and Computer-assisted Analyses—The BamHI inserts from pAGB-1, an EcoRI-BamHI restriction fragment of...
Full-length cDNA Encoding Human α-N-Acetylgalactosaminidase

**Fig. 2.** Nucleotide and predicted amino acid sequences of the 2.2-kb pAGB-3 and 3.6-kb pAGB-35 cDNA inserts containing the complete coding region for human α-GalNAc. The pAGB-3 and pAGB-35 tryptic peptides (r) of the purified enzyme. CHO indicates potential sites of N-glycosylation. Overlines indicate colinear amino acid sequence obtained by microsequencing the N-terminal (N-Z’er) and tryptic peptides (r) of the purified enzyme. CHO indicates potential sites of N-glycosylation.

---

**Note:** The DNA and amino acid sequences were obtained using the GCG program package (28). The DNA sequences were analyzed using the Wisconsin Genetics Computer Group software (28). The amino acid sequences were derived from the DNA sequences using the Wisconsin Genetics Computer Group software (28). The authors thank Drs. J. L. Allred and J. R. B. Lewis for the synthesis of the oligonucleotides and for helpful discussions. This work was supported by National Institutes of Health grant AI-14807 (M. T. J. S.), the National Science Foundation (E. M. T.), and the Swedish Medical Research Council (J. M.).
calcium-phosphate precipitation (31). Cells were harvested at 24-h intervals after transfection and assayed for \(\alpha\)-GalNAc activity as previously described (18). One unit of enzymatic activity is equal to that amount of enzyme required to hydrolyze 1 nmol of 4-MU-\(\alpha\)-GalNAc/h. Protein concentrations were determined by the fluorescamine method (21).

**Northern Hybridization and Cap Site Analyses—**Total RNA was isolated from human lymphoblasts, fibroblasts, and placenta, and Northern hybridization was performed using the nick-translated pAGB-3 insert as probe (26). Alternatively, the pAGB-3 insert was subcloned into pGEM-4Z (Promega, Madison, WI), and radiolabeled \(\alpha\)-GalNAc riboprobe, rAGB-3, was generated using the Promega riboprobe system and used for Northern hybridization. For identification of the \(\alpha\)-GalNAc cap site, two unique, overlapping 30-mer oligonucleotide primers were synthesized corresponding to regions 60 and 75 bp from the 5'-end of the pAGB-3 cDNA and end-labeled with \(\alpha\)-[\(\beta\]-32P]ATP. The RNA/DNA heteroduplexes were isolated as described (24, 26). cDNA was synthesized from 10 \(\mu\)g of total RNA or 2.5 \(\mu\)g of brain poly(A)\(^+\) mRNA (Clontech, Palo Alto, CA) using the BRL cDNA Synthesis Kit. Bacteriophage DNA (0.1 \(\mu\)g) was radiolabeled with 100 \(\mu\)Ci of \([\beta\]S\]methionine (Amersham Corp.) per 100 \(\mu\)l of loading dye (0.3% xylene cyanol, 0.3% bromphenol blue, 0.37% garencine, pH 7.0, in formamide). The RNA/DNA heteroduplexes were denatured at 68 °C for 3 min, and an aliquot was electrophoresed on a standard 8 M urea, 5% polyacrylamide sequencing gel.

**Construction of p91-\(\alpha\)-GalA/\(\alpha\)-GalNAc7 and Characterization of the Hybrid Protein—**A plasmid containing \(\alpha\)-Gal A exons 1 through 6 from pcDAG-126 (32) was fused to the 3' region of pAGB-3 \(\alpha\)-GalNAc insert which corresponded in position to \(\alpha\)-Gal A exon 7. The hybrid cDNA, designated \(\alpha\)-GalA/\(\alpha\)-GalNAc7 was constructed with the sense and antisense primers indicated above using a PCR-based method (33) and sequenced. The p91-\(\alpha\)-GalA/\(\alpha\)-GalNAc7 insert was subcloned into the expression vector, p91-\(\alpha\)-AGA (34), which contained the entire \(\alpha\)-Gal A cDNA, also was transiently expressed. The \(\alpha\)-Gal A and \(\alpha\)-GalNAc enzymatic activities and enzyme proteins were detected with 4-MU substrates and by immunoblotting or immunoprecipitation with the respective polyclonal antibodies as described above. For immunoprecipitation studies, transfected COS-1 cells were radiolabeled with 100 \(\mu\)Ci of \([\beta\]S\]methionine (Amersham Corp.) per 100 \(\mu\)l at 48-h post-transfection. The cells were harvested at 72 h after transfection and immunoprecipitated as described (26) with Staphylococcus aureus cells (Pansorbin, Calbiochem, San Diego, CA).

**Analysis of cDNA and Genomic Sequences by PCR Amplification and Primer Extension—**For PCR amplification of the putative alternatively spliced region, the 30-mer sense and antisense primers (described above) were used to amplify the 1) reverse-transcribed mRNA from various human sources, 2) cDNA inserts from clones pAGB-4 to 34, and 3) the gagB-1 genomic sequence. DNAs from pAGB-4 to 34 cDNA clones and the gagB-1 genomic clones were isolated as described (24, 26). cDNA was synthesized from 10 \(\mu\)g of lymphoblast, fibroblast, and placental total RNA or 2.5 \(\mu\)g of brain poly(A)\(^+\) mRNA (Clontech, Palo Alto, CA) using the BRL cDNA Synthesis Kit. Bacteriophage DNA (0.1 \(\mu\)g), reverse-transcribed mRNA (0.1 \(\mu\)g) or genomic cosmid DNA (0.1 \(\mu\)g) was PCR-amplified using 20 \(\mu\)M of each primer and the GeneAmp DNA Amplification Reagent Kit (Perkin-Elmer Cetus, Norwalk, CT). Each PCR cycle consisted of 1 min of denaturation at 94 °C, 2 min of annealing at 54 °C, 2 min of annealing at 37 °C, and a 7-min extension at 60 °C. The PCR products were phenol extracted, ethanol precipitated, and resuspended in 20 \(\mu\)l of H\(_2\)O. An aliquot (2 \(\mu\)l) of each PCR reaction was analyzed by electrophoresis on agarose gel using HindIII-digested lambda and HaeIII digested phiX174 DNAs as size standards. For identification of potential stops during reverse transcription of the region surrounding the pcd-HS1204 insertion, a unique 32-mer, 5'-AGTAGTAAGCTTTCATATACGACCCGGT-3', was used to extend 10 \(\mu\)g of total placental RNA or 1 \(\mu\)g of rAGB-3 generated in vitro by the Promega riboprobe system as described above.

**RESULTS AND DISCUSSION**

**Purification and Characterization of Human \(\alpha\)-GalNAc—**Human \(\alpha\)-GalNAc was purified to homogeneity (specific activity = -370,000 units/mg protein) as assessed by the presence of only the 48- and 117-kDa species on NaDodSO\(_4\)/PAGE (Fig. 1, inset). The 117-kDa species was not reduced by boiling or by dialysis against 8 M urea in the presence of \(\beta\)-mercaptoethanol. The 27 microsequenced N-terminal residues of the electroeluted 117-kDa species were identical to those of the 48-kDa species. Further evidence that the 117-
kDa species was a homodimer of the 48-kDa glycoprotein subunit. The finding that the tryptic digests (and chymotryptic digests, not shown) of both species had essentially identical HPLC profiles (Fig. 1). Microsequencing of the N terminus and seven tryptic peptides from the 48-kDa species identified a total of 129 non-overlapping a-GalNAc residues. For library screenings, synthetic oligonucleotide mixtures (17- to 26-mers) were constructed to contain all possible codons for selected amino acid sequences from the N terminus and three internal tryptic peptides (Figs. 1 and 2).

Isolation, Characterization, and Expression of a Full-length cDNA—Screening of 2 × 10^6 recombinants from the pcD human fibroblast cDNA library with a 26-mer oligonucleotide mixture of 576 species corresponding to internal peptide T-106A detected two putative positive clones. pAGB-1, which hybridized with all four oligonucleotide mixtures, had a 1.8-kb insert with an open reading frame of 1242 bp, a 514-bp 3' untranslated region, and a poly(A) tract, but no apparent 5' untranslated sequence. Authenticity was established by collinearity of the pAGB-1 insert's predicted amino acid sequence with 129 microsequenced residues of the purified protein. In order to isolate a full-length cDNA, the 0.9-kb 5'-BamHI fragment from the pAGB-1 insert was radiolabeled and used to screen a human placental cDNA library. Of 92 putative positive clones (pAGB-3 to 34), pAGB-3 contained the longest insert and was sequenced in both orientations. As shown in Fig. 2, the 2158-bp pAGB-3 insert had a 344-bp 5'-untranslated region, a 1236-bp open reading frame which encoded 411 amino acids, a 514-bp 3'-untranslated region and a 64-bp poly(A) tract. An upstream, inframe ATG occurred at -192 nt, but there were inframe termination codons at -141, -135, and -120 nt, indicating that the -192 ATG was nonfunctional. A single consensus polyadenylation signal (AATAAA) at nt 3100-3105 (Fig. 2) downstream from the signal at nt 1729-1734 in the pAGB-3 cDNA indicated that alternative polyadenylation was responsible for the generation of the larger transcript. The occurrence of the 3.6 kb transcript indicates that the a-GalNAc gene has at least another transcriptional start site upstream from the -347 cap site of the 2.2-kb transcript.

Sequence Homology between a-GalNAc with a-Gal A—Computer-assisted searches of nucleic acid and protein data bases revealed no significant amino acid sequence similarities between a-GalNAc and that of any other DNA or protein sequence except for human a-Gal A (32). Comparison of the nucleic acid and deduced amino acid sequences of the full-length a-GalNAc and a-Gal A cDNAs revealed 55.8 and 46.5% overall homology, respectively. Since the intron/exon junctions and the entire genomic sequence encoding human a-Gal A have been determined (32, 36), it was possible to compare the a-GalNAc amino acid sequence with that deduced from each of the seven a-Gal A exons (Fig. 4). Notably, there was remarkable identity (56.4%) between the a-GalNAc sequences corresponding to those of a-Gal A exons 1 through 6. For example, all 8 cysteine residues in a-GalNAc were present in a-Gal A, 10 and 20 were conserved in identical positions in a-GalNAc, respectively. Furthermore, all four of the a-Gal A N-glycosylation sites were conserved in a-GalNAc. Putative functional domains were suggested by shorter stretches of amino acid homology shared by a-GalNAc, a-Gal A, yeast a-galactosidase (Mel 1) (38) and/or Escherichia coli a-galactosidase (Mel A) (39) in a-Gal A exons 1 through 6. In contrast, there was little, if any, similarity in the predicted a-GalNAc carboxyl-terminal amino acid se-

![Fig. 5. Immunoprecipitation of human a-GalNAc, a-Gal A, and a hybrid a-Gal A/a-GalNAc protein expressed in COS-1 cells. Lane 1, mock transfection; lane 2, p91-a-Gal A transfection; lanes 3 and 4, two independent constructs of p-a-GalA6/a-GalNAc7 transfections; lane 5, p91-AGB-3 transfection. Lanes 1-4 were immunoprecipitated with rabbit anti-human a-Gal A, and lane 5 was immunoprecipitated with rabbit anti-human a-GalNAc.](http://www.jbc.org/content/218/52/21863.full.html)

Downloaded from http://www.jbc.org/ on March 24, 2010
Full-length cDNA Encoding Human α-N-Acetylgalactosaminidase

Fig. 6. Partial genomic sequence of human α-GalNAc including an intron between coding nt 957 and 958. A, rAGB-3: partial α-GalNAc RNA sequence (nt 909–969) corresponding to the 3′-end of α-Gal A exon 6 (32). The indicated stem and loop structure between nt 918 and 937 had a AG of -11.6 (29). The overlapped antisense and sense sequences shown in bold are inverted and direct repeats derived from nt 919–967 of pAGB-3 that are in the 70-bp insertion in pcD-HS1204 (20). The 45-bp deletion in clone pAGB-13 is indicated in italics. B, the genomic α-GalNAc sequence from coding nt 790–1053 (supercodon) includes a 1754-nucleotide intron between nt 959 and 958 which corresponds in position to the α-Gal A exon 6 and 7 boundaries. Dashed lines, the 5′ splice donor sequence; solid underlines, putative branch point sequences; dotted underlines, putative polypyrimidine tracts at the 3′ acceptor sites for the normal gene and mutant pAGB-13; and asterisks, differences from the consensus sequence (44).
tion of an enzyme-substrate complex was prevented by steric hinderance. Future mutagenesis studies of this region in \( \alpha \)-Gal A and \( \alpha \)-GalNAc may provide further information concerning these possibilities.

The finding of extensive homology between \( \alpha \)-GalNAc and \( \alpha \)-Gal A suggested that they evolved by duplication and divergence of an ancestral sequence for \( \alpha \)-Gal A exons 1 through 6. Although there is little, if any, homology among the other lysosomal amino acid sequences (i.e., non-lysosomal domains), there are notable examples of lysosomal enzyme subunits, pseudogenes, or gene families which presumably evolved by duplication and divergence (e.g., 40–42). Future comparison of the \( \alpha \)-GalNAc and \( \alpha \)-Gal A intron/exon boundaries should provide further information on the evolution of these lysosomal genes which encode structurally related, but functionally specific glycohydrolases.

**Primer Extension and PCR and Sequence Analyses of cDNA and Genomic Sequences**—During the course of these studies, Tsuji et al. (29) reported a similar human \( \alpha \)-GalNAc cDNA sequence which differed from pAGB-3 by a 70-bp insertion after nt 957 (Fig. 6A) and by several substitutions (nt 493, 494, 542, 614, and 667). The 70-bp insertion consisted of three inverted repeats (nt 919–926, 919–936, and 919–944) and a direct repeat (nt 940–957) from the pAGB-3 coding sequence nt 919–957. Analysis of the pAGB-3 cDNA sequence from nt 760–1053 using an RNA-folding program (29) predicted a stem and loop structure from nt 918 to 937 (Fig. 6A) which could stall or stop reverse transcription of the \( \alpha \)-GalNAc mRNA during cDNA synthesis. To determine if this secondary structure could cause cDNA synthesis errors in library construction, a 32-mer oligonucleotide primer was used to extend total placental RNA and \( \alpha \)-GalNAc transcripts generated in vitro with the riboprobe construct, rAGB-3. Stops of varying intensity were observed from nt 903 and 1009, including two weak stops at the 3' base (nt 940) and 5'-end (nt 921) of the stem and loop structure (Fig. 6A). However, there were no strong stops in this region. Although the actual mechanism is unknown, these findings were consistent with the 70-bp insertion resulting from a complex abnormality involving an RNA-DNA duplex in cDNA library construction (43). Another possibility would be an insertion due to a complex strand-switching event involving DNA polymerase I (45).

Alternatively, this 70-bp insertion may have resulted from alternative splicing, although the insertion predicts a truncated \( \alpha \)-GalNAc polypeptide of 358 residues. To investigate the possible occurrence of \( \alpha \)-GalNAc transcripts with a 70-bp insertion after pAGB-3 nt 957, PCR was used to amplify this region in 1) reverse-transcribed mRNA from various sources, 2) the cDNA inserts from clones pAGB-4 to 34, and 3) the pAGB-1 genomic clone. If the cDNA inserts or reverse-transcribed RNAs contained the 70-bp insert, a 290-bp PCR product would be observed, whereas the absence of the insert would result in a 220-bp PCR product. Only the 220-bp product was observed in PCR-amplified reverse-transcribed total RNA from human lymphoblasts, fibroblasts, and placenta, or in poly(A)+ mRNA from brain (not shown). Thus, these analyses did not detect longer or shorter transcripts. All of the pAGB-4 through 34 cDNA inserts had only the 220-bp PCR product with the exception of pAGB-13, which had an inframe 45-bp deletion after pAGB-3 nt 957 (i.e., deleted nt 958–993). A short direct repeat (ACAAG) was present at both breakpoint junctions. Notably, the deletion occurred at the identical 5' site of the 70-bp insertion in pcd-HS1204 (14) (Fig. 6A).

Subsequent sequencing of the region including pAGB-3 codons 254–351 in the genomic clone, pAGB-1, revealed a 2046-bp sequence containing a 1754-bp intron between pAGB-3 nt 957 and 958. The intronic sequence had no homology with \( \alpha \)-Gal A intron 6, contained two Alu-repetitive sequences in reverse orientation and did not have the 70-bp insertion in either orientation (Fig. 6B). It was remarkable that both the pAGB-13 deletion and the pcd-HS1204 insertion occurred at the 5' donor splice site of this intron (i.e., after exonic nt 957). Perhaps the location of the conserved lariat branch point sequences in the intron far upstream (94 and 199 bp) from the 3' splice site may impair splicing (44). This concept is supported by the pAGB-13 deletion in which the more closely positioned cryptic lariat branch point and 3' splice site were used. Thus, this intron or surrounding region may have a unique sequence and/or secondary structure that impedes the fidelity of hnRNA processing. Since the intron/exon junction after coding nt 957 also is the site of divergence between the \( \alpha \)-Gal A and \( \alpha \)-GalNAc sequences, this region also may be mechanistically important in the evolution of human \( \alpha \)-GalA.

In conclusion, the availability of an authentic full-length cDNA encoding human \( \alpha \)-GalNAc should permit the characterization of the structure/function and evolutionary relationships of \( \alpha \)-GalNAc and \( \alpha \)-Gal A as well as the identification of the molecular lesions that cause Schindler disease.

**REFERENCES**

1. Kint, J. A. (1971) *Arch. Int. Physiol. Biochem.* 79, 633–644
2. Beutler, E., and Kuhl, W. (1972) *Am. J. Hum. Genet.* 24, 237–249
3. Romeo, G., Childs, B., and Mizeon, B. R. (1972) *FEBS Lett.* 27, 161–166
4. Wood, S., and Nadler, H. L. (1972) *Am. J. Hum. Genet.* 24, 250–255
5. Ho, M. W., Beutler, S., Tennant, L., and O'Brien, J. (1972) *Am. J. Hum. Genet.* 24, 256–266
6. Deansick, R. J., Allen, K. Y., Desnick, E. J., Raman, M. K., Bernholz, R. W., and Krivit, W. (1973) *J. Lab. Clin. Med.* 81, 157–171
7. Deansick, R. J., Jr., and Bishop, D. F. (1989) in *The Metabolic Basis of Inherited Disease* ( Stellar, C. R., Beaudet, A. L., Sly, W. S., and Vallee, D., eds) pp. 1751–1796, McGraw Hill, New York
8. Beutler, E., and Kuhl, W. (1972) *J. Biol. Chem.* 247, 7195–7200
9. Callahan, J. W., Lasilla, E. L., Den Tandt, W., and Philippart, M. (1973) *Biochem. Med.* 7, 424–431
10. Dean, K. J., Sung, S., and Sweeney, C. C. (1977) *Biochem. Biophys. Res. Commun.* 77, 1411–1417
11. Schram, A. W., Hamwe, M. N., and Tager, J. M. (1977) *Biochim. Biophys. Acta* 482, 138–144
12. Kusik, J. W., Quirk, J. M., and Brady, R. O. (1978) *J. Biol. Chem.* 253, 104–100
13. Dean, K. J., and Sweeney, C. C. (1979) *J. Biol. Chem.* 254, 10001–10005
14. Bishop, D. F., Dean, K. J., Sweeney, C. C., and Desnick, R. J. (1980) in *Enzyme Therapy in Genetic Disease 2* (Desnick, R. J., ed.), pp. 17–32, Alan R. Liss, Inc., New York
15. deGroot, P. G., Westerveld, A., Meera-Khan, P., and Tager, J. M. (1978) *Hum. Genet.* 44, 305–312
16. Sweeney, C. C., LeDonne, J. C., and Robbins, P. W. (1983) *Arch. Biochem. Biophys.* 223, 158–165
17. van Diggelen, O. P., Schindler, D., Willemsen, R., Boer, M., Kneef, W. J., Huizjna, J. G. M., Blom, W., and Coljaard, H. (1988) *J. Inherited Metab. Dis.* 11, 349–357
18. Schindler, D., Bishop, D. F., Wolfe, D. E., Wang, A. M., Egge, H., Lemeuc, R. U., and Desnick, R. J. (1988) *N. Engl. J. Med.* 320, 1735–1740
19. Schindler, D., Kanaki, T., and Desnick, R. J. (1990) *Clin. Chim. Acta* 190, 89–97
20. Tsuji, S., Yamashita, T., Hiroya, M., Isobe, T., Okuyama, T., Sakimura, K., Takahashi, Y., Mishizawa, M., Uda, Y., and Miyatake, T. (1989) *Biochem. Biophys. Res. Commun.* 163, 1498–1504
21. Bishop, D. F., and Desnick, R. J. (1981) J. Biol. Chem. 256, 1307-1316
22. Calhoun, D., Bishop, D. F., Bernstein, H. S., Quinn, M., Hantzopoulos, P., and Desnick, R. J. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 7364-7368
23. Hunkapillar, M. W., Lujan, E., Ostrander, F., and Hood, L. E. (1975) Methods Enzymol. 31, 227-236
24. Tsai, S. F., Bishop, D. F., and Desnick, R. J. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 7049-7053
25. Hunkapillar, M. W., and Hood, L. E. (1983) Science 219, 650-659
26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (eds) (1989) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY
27. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980) J. Mol. Biol. 143, 161-178
28. Wolf, H., Modrow, S., Motz, M., Jameson, B. A., Hermann, G., and Fortsch, R. (1988) Comput. Appl. Biosci. 4, 187-191
29. Zoller, M. (1989) Methods Enzymol. 180, 262-288
30. Wong, G. G., Wittek, J. S., Temple, P. A., Wilkins, K. M., Leary, A. C., Luxenberg, D. P., Jones, S. S., Brown, E. L., Kay, R. M., Orr, F. C., Shoemaker, C. S., Golde, D. W., Kaufman, R. J., Hewick, R. M., Wang, E. A., and Clark, S. C. (1985) Science 228, 810-813
31. Chen, C., and Okayama, H. (1987) Mol. Cell. Biol. 7, 2745-2752
32. Bishop, D. F., Kornreich, R., and Desnick, R. J. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 3903-3907
33. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Gene (Amst.) 77, 51-59
34. Bishop, D. F., Kornreich, R., Eng, C. E., Ioannou, Y. A., Fitzmaurice, T. F., and Desnick, R. J. (1988) Lipid Storage Disorders: Biological and Medical Aspects, pp. 809-822, Plenum Publishing Co., New York
35. Berget, S. M. (1984) Nature 309, 179-182
36. von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690
37. Kornreich, R., Desnick, R. J., and Bishop, D. P. (1989) Nucleic Acids Res. 17, 3301-3302
38. Liljestrom, P. L. (1985) Nucleic Acids Res. 13, 7257-7268
39. Liljestrom, P. L., and Liljestrom, P. (1987) Nucleic Acids Res. 15, 2213-2220
40. Proia, R. L. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 1883-1887
41. Horowitz, M., Wilder, S., Horowitz, A., Keiner, O., Gelbart, T., and Beutler, E. (1989) Genomics 4, 87-96
42. Schuchman, E. H., Jackson, C. E., and Desnick, R. J. (1990) Genomics 6, 149-158
43. Roberts, J. D., Preston, B. D., Johnston, L. A., Soni, A., Loeb, L. A., and Kunkel, T. A. (1989) Mol. Cell. Biol. 9, 468-476
44. Papanicolaou, C., and Ripley, L. S. (1989) J. Mol. Biol. 207, 335-353
45. Reed, R., and Maniatis, T. (1988) Genes & Dev. 2, 1968-1976
Human alpha-N-acetylgalactosaminidase-molecular cloning, nucleotide sequence, and expression of a full-length cDNA. Homology with human alpha-galactosidase A suggests evolution from a common ancestral gene.

A M Wang, D F Bishop and R J Desnick

J. Biol. Chem. 1990, 265:21859-21866.