EXPERIMENTAL and MOLECULAR MEDICINE, Vol. 32, No. 4, 187-192, December 2000

Characterization of an unusual variant mRNA of human lysosomal α-mannosidase

Huaichang Sun[1,2] and John H. Wolfe[1,3]

[1] Laboratory of Pathology and Center for Comparative Medical Genetics, School of Veterinary Medicine, University of Pennsylvania, 3800 Spruce Street, Philadelphia, PA 19104
[2] Present address: Department of Veterinary Medicine, College of Animal Science and Veterinary Medicine, Yangzhou University, Yangzhou, Jiangsu 225009, P. R. China
[3] Corresponding author: Tel, +1-215-590-7028; Fax, +1-215-590-3779; E-mail, jhwolfe@vet.upenn.edu

Accepted 14 November 2000

Abstract

Lysosomal α-mannosidase (EC 3.2.1.24) is an exoglycosidase in the glycoprotein degradation pathway and is encoded by a 3.0 kb cDNA. A 2.3 kb cDNA from a minor species of HeLa cell mRNA was discovered by RT-PCR cloning. Southern blotting and PCR analysis of the HeLa cell genomic DNA showed that the 2.3 kb message was encoded by the lysosomal α-mannosidase gene. Sequence comparison of the cDNA with the corresponding genomic DNA indicated that the 2.3 kb message was generated by an unusual intra-exonic joining event.

Keywords: lysosomal α-mannosidase, intron sequences, mRNA, splice signal sequence

Introduction

Lysosomal α-mannosidase is an exoglycosidase in the glycoprotein degradation pathway (Daniel *et al*., 1994). A deficiency of this enzyme causes a lysosomal storage disease, α-mannosidosis (Thomas and Beaudet, 1995). The human enzyme gene has been mapped to chromosome 19p13.2-q12 by analysis of human and rodent somatic cell hybrid mapping panels (Kaneda *et al*., 1987) and its genomic structure has been shown to span 21.5 kb with 24 exons (Riise *et al*., 1997). Several human lysosomal α-mannosidase cDNA sequences have been reported (Nebes and Schmidt, 1994; Liao *et al*., 1996; Nilssen *et al*., 1997; Sun *et al*., 1999), but only one has been shown to express high levels of α-mannosidase activity in α-mannosidosis cells (Sun *et al*., 1999). Splicing of eukaryotic pre-mRNAs involves the accurate selection of the correct 5' and 3' splicesites (Krainer and Maniatis, 1988; Green, 1991). The consensus for the 5' and 3' splice site sequences in higher eukaryotes has been determined (Shapiro and Senapathy, 1987). Generally, sequences that show a better match to the consensus are more frequently used as authentic or strong splice sites, whereas the sites with a poorer match to the consensus tend to be inactive or weak (Fu *et al*., 1988; Lowery and Van Ness, 1988; Brunak and Engelbrecht, 1991). However, these consensus sequences are not sufficient to account for the observed high specificity of splice site selection. Other sequence elements, such as the downstream exons and the secondary structures of the region around splice sites, are also involved in the selection of splice sites (Solnic, 1985; Watakabe *et al*., 1992). Alternative splicing is a powerful post-transcriptional mechanism for regulating gene expression. By utilizing different splice sites within the same precursor RNA, multiple transcripts and the variant proteins can be generated from a single gene (Smith *et al*., 1989). Here we present evidence that an unusual transcript of human lysosomal α-mannosidase can be generated in HeLa cells by splicing at non-canonical sequences within 2 exons.

Materials and Methods

RT-PCR

Total RNA was prepared from HeLa cells using RNeasy Kit (Qiagen) according to the manufacturer's instructions. 5 μg of the RNA was used to synthesize the first strand cDNA using Superscript™ Preamplification System for First Strand cDNA Synthesis (Gibco BRL). The reaction was performed at 42°C for 50 min using an oligo (dT) primer. 4 μl of the first strand cDNA was used to amplify the double strand cDNA using Expand™ High Fidelity PCR System (Boehringer Mannheim) and a touch down program of 30 cycles, with denaturation at 94°C for 30 s (4 min for the first cycle), annealing at 62°C for 30 s, and elongation at 68°C for 3-7 min. The primers were designed according to the previously reported retina/muscle composite cDNA for human lysosomal α-mannosidase (Nebes and Schmidt, 1994) and are shown in Figure 1A. The 3' end of the cDNA was extended using the 3' RACE System (Gibco BRL) according to the manufacturer's instructions.

Preparation of genomic DNA

HeLa cell genomic DNA was prepared using IsoQuick

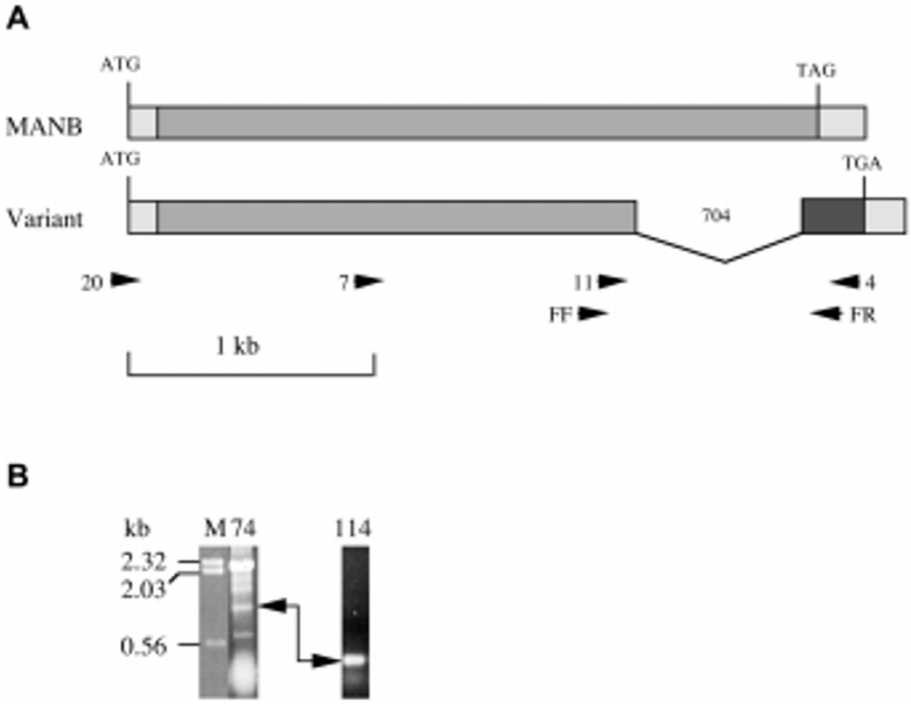

Figure 1. Human lysosomal α-mannosidase cDNA and 2.3 kb variant. A. The light grey box at the 5' end represents the leader sequence, the long medium grey boxes represent the common coding regions, and the dark grey box represents the altered coding region at the 3' end of the 2.3 kb variant. The gap shows the location of the 704 bp deletion region. Arrows indicate the locations of the primers used: primers 20 and 4 were used for amplification of the full-length cDNAs; primers 7, 11 and 4 were used for nested PCRs; and primers FF and FR were used to amplify the 3.5 kb genomic DNA. B. Confirmation of the 2.3 kb message by sequential nested PCRs. HeLa cell RNA was amplified by RT-PCR using a nested primer man 7 and primer man 4 (74), of which products were ranged from 1.3 kb (corresponding to the 2.3 kb variant) to 2.0 kb (corresponding to 3.0 kb α-mannosidase message).The 1.3 kb band was purified and reamplified by PCR using another nested primer man 11 and primer man 4 (114). Arrows show the PCR products corresponding to the 2.3 kb minor message. M: molecular weight marker.

Nucleic Acid Extraction Kit (MicroProbe Corporation). The contaminant RNA was removed by RNase digestion, and the resulting DNA was extracted once with the Buffer 2 and precipitated with 2 volumes of ethanol.

PCR and Nested PCR

The target genomic DNA or RT-PCR fragments purified from agarose gels was amplified using the Expand Long Template PCR System (Boehringer Mannheim) with 500 μM dNTPs, 300 nM primers, 300 ng DNA, Buffer 3 (containing 20% DMSO and 1% Tween 20), and the above program for RT-PCR. After 25 cycles of amplification, 10 μl of each PCR reaction was analyzed by agarose gel electrophoresis. The primers used for amplifying target genomic DNA or for nested PCR are indicated in Figures 1A and 3C.

DNA sequencing

The "full-length" RT-PCR product was cloned in pCRII vector (Invitrogen). Three separate reactions of other PCRs were purified using Qiagen's Gel Extraction Kit and pooled to dilute out random errors. The DNAs were sequenced on both strands on an Applied Biosystem 373A DNA Sequencer using a dye-terminator method.

Southern blotting

The purified HeLa cell genomic DNA was digested with different restriction enzymes and separated by 1.0% agarose gel electrophoresis before transfer onto Hybond-N Membrane (Amersham Life Science). After UV cross-linking at 312 nm for 4 min, the membrane containing the genomic DNA was probed with human α-mannosidase cDNA (Nebes and Schmidt, 1994) labeled with α-[^{32}P] dCTP using Ready-To-Go^{TM} DNA Labeling Beads (Pharmacia Biotech). The hybridization was performed according to a standard method at high stringency (Ausubel *et al.*, 1994).

Results

Human lysosomal α-mannosidase cDNA was amplified by RT-PCR of HeLa cell mRNA and the PCR product was cloned in pCR II vector. After transformation of competent cells, white colonies were picked for preparation of plasmid DNA and restriction enzyme analysis. Six clones containing inserts ranging from 2.3 to 3.0 kb were evaluated. Both the 2.3 and 3.0 kb clones were sequenced and found to be identical except for a

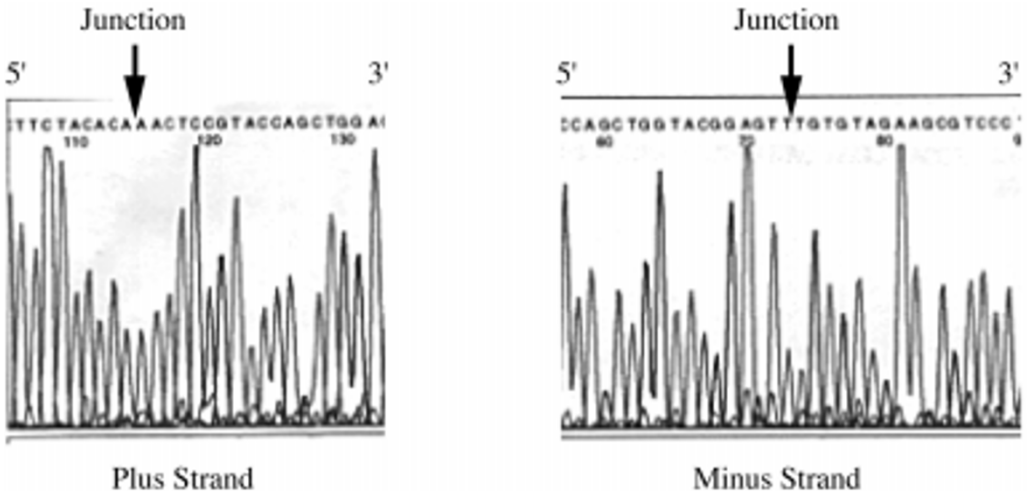

Figure 2. Partial sequences of the secondary nested PCR product. Three separate secondary nested PCR reactions (using primers 11 - 4 in Figure 1A) were pooled and sequenced in both directions (plus strand and minus strand). Arrows indicate the junction site where the 704 bp deletion occurred.

704 bp deletion in the smaller cDNA near the 3'- end of the coding region (Figure 1A). The deletion caused a frameshift in the smaller cDNA and resulted in an alternate TGA stop codon 59 bp downstream of the predicted TAG stop codon in the 3.0 kb cDNA (Figure 1A). The 3.0 kb cDNA has been shown to encode high activity human lysosomal α-mannosidase (Sun *et al.*, 1999). The 2.3 kb cDNA was predicted to contain an open reading frame (ORF) of 2325 bp and to encode a mature protein of 84 kD with 7 potential N-linked glycosylation sites after cleavage of the predicted 26 amino acid signal peptide.

To exclude that the 2.3 kb clone was an anomaly generated by PCR and/or cloning errors, additional RT-PCR reactions were performed using a set of nested primers (Figure 1A). Agarose gel electrophoresis revealed an expected 1.6 kb product (Figure 1B), which was purified and reamplified using another set of nested primers (Figure 1A). Once again, a product of the expected size (0.2 kb) was obtained (Figure 1B). Sequence analysis showed that the secondary nested PCR product matched exactly to the original 2.3 kb cDNA (Figure 2).

To study the relationship between the 2.3 kb and 3.0 kb cDNAs at the genome level, HeLa cell genomic DNA was digested with different restriction enzymes and probed with human lysosomal α-mannosidase cDNA (Nebes and Schmidt, 1994). More than one band was present in the *Bam* HI, *Eco* RI and *Hind* III digests, whereas only one band of about 20 kb was detected in the *Bgl* II digest (Figure 3A), indicating that the two cDNAs were derived from a single lysosomal α-mannosidase gene. To confirm this, HeLa cell genomic DNA was amplified by PCR using primers flanking the 704 bp deletion region in the smaller cDNA (Figure 1A). Once again only one band of 3.5 kb was revealed by agarose gel electrophoresis of the PCR product (Figure 3B).

Sequence analysis showed that the 3.5 kb genomic DNA spanned 8 exons (exons 17-24, Riise *et al.*, 1997) and that the missing region in the 2.3 kb cDNA corresponded to sequences in exons 18-24, skipping 5 exons in the 3.0 kb lysosomal α-mannosidase cDNA. This exactly accounted for the 704 bp deletion (Figure 3C). All of the boundaries of the 5 exons conformed to the GT-AG rule for nuclear pre-mRNA splicing, as described previously (Riise *et al.*, 1997). In contrast, the junctions where the 704 bp deletion occurred were not flanked by either a consensus splice donor or acceptor (Figure 4A). A consensus splice donor site was not found until 33 bp downstream of the 5'-end of the deleted sequence (Figure 4A), and a splice acceptor site was present 17 bp upstream of the 3'-end of the deleted sequence (Figure 4A). A CA repeat was present at the junction site in the 2.3 kb cDNA (Figure 4A). Since there was a CACA in the 18th exon and a CA in the 24th exon at the junction sites, the second A or CA could be from either exon. In addition, two Alu sequences (Ya-5 and Sq subfamilies) were present in the 18th and 20th introns in opposite orientation (big arrows in Figure 3C). Computer analysis for homologies (>70%) identified a pair of 30 bp sequences near the 5'- and 3'-junction sites (small arrows in Figure 3C). A 19 bp segment of this contained 16 complementary pairs, of which 13 were G-C (Figure 4B).

Discussion

There are a number of different possibilities explaining these results. First of all, PCR is still notorious for generating random mutations, even though the error rate is presently fairly low (1/5000-6000) due to use of high fidelity DNA Taq polymerase. However, it is very unlikely

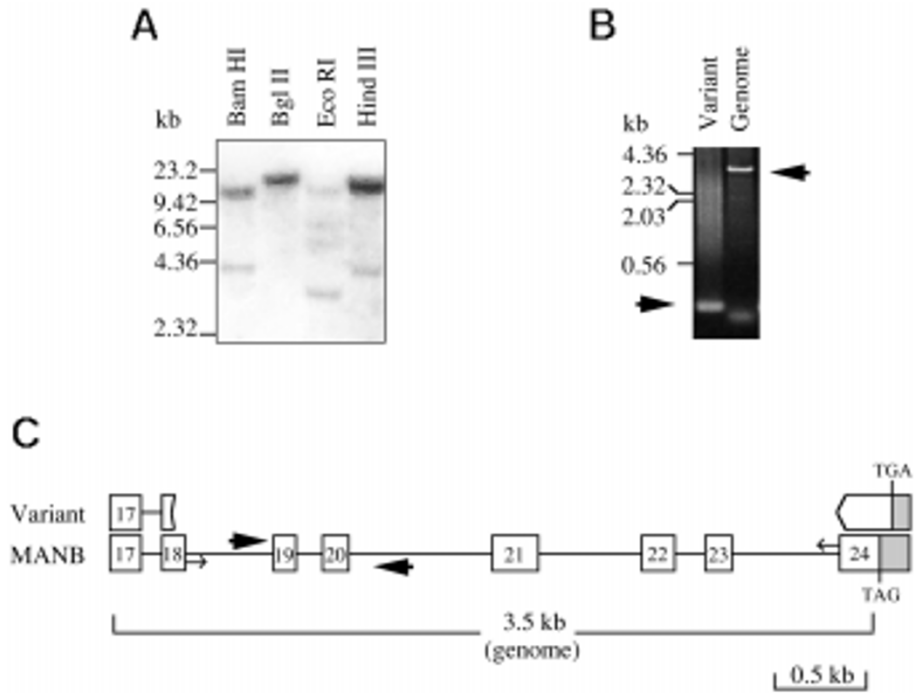

Figure 3. Characterization of genomic DNA in the region corresponding to the deleted portion of the 2.3 kb cDNA. A. Southern blot hybridization. HeLa cell genomic DNA was digested with restriction enzymes *Bam* HI, *Bgl* II, *Eco* RI or *Hind* III, and probed with a full-length cDNA for α-mannosidase at high stringency. B. Amplification of potential fusion exon(s) by PCR. A plasmid containing the 2.3 kb cDNA (variant) and HeLa cell genomic DNA (genome) were amplified using primers flanking the junction site. Arrows point to the variant cDNA and the genomic fragment. Primer dimers can be seen at the bottom of the genomic lane. C. The intronic/exonic structures of the 3.5 kb genomic DNA corresponding to the 704 bp deletion in the 2.3 kb cDNA. Open boxes, lines, shaded boxes, bold arrows, and small arrows represent exons, introns, untranslated regions, Alu sequences, and small repeats, respectively.

that the 704 bp deletion in the 2.3 kb cDNA was caused by random errors introduced by reverse transcriptase and/or Taq DNA polymerase since we found exactly the same sequence junction in independent experiments. Another potential PCR artifact is the so-called "jumping PCR", which requires stable complementary structures to bring two distant sequences together. However, the 2.3 kb cDNA was reproducibly amplified when two sequential nested primers were used and DMSO was included in the PCR buffer (see 'Materials and Methods'), which would destroy potential secondary structures in the template DNA. Furthermore, computer analysis did not find similar complementary sequences in the 3.0 kb cDNA, which could allow the DNA Taq polymerase to jump 704 bp to generate the 2.3 kb cDNA. Although a pair of 30 bp complementary sequences were found in the 3.5 kb genomic DNA (which would be present in the pre-mRNA) corresponding to the deleted cDNA sequence, they were about 20 and 15 bp away from the 5'- and 3'- junction site, respectively (Figure 4B). Thus these elements would not appear to explain the precise junction sequence found in the smaller cDNA.

Another explanation would be the presence of more than one copy of the gene, with one of them bearing the deletion and one with the normal sequence. This could occur if duplication occurred to create a new gene or if a pseudogene was present which is illegitimately transcribed. However, in these situations, analysis of genomic DNA should reveal differences in sequence. Southern blotting of HeLa cell genomic DNA, using the complete human lysosomal α-mannosidase cDNA as the probe, revealed only one band of about 20 kb in the *Bgl* II digest. This indicates that the two cDNAs were derived from a single gene. Previous experiments also found evidence for only a single gene for the human lysosomal α-mannosidase, which is 21.5 kb with 24 exons (Riise *et al.*, 1997). This was confirmed by our PCR analysis of the HeLa cell genomic DNA using primers flanking the deletion region of the smaller cDNA, in which no sequence accounting for the 704 bp difference was found (Figure 3B).

The data thus indicate that the 2.3 kb message was generated from the lysosomal α-mannosidase gene by alternative splicing. In this case, the pair of 30 bp sequences with 16 complementary pairs, especially the 13 G-C pairs flanking the deletion region, might form a strong secondary structure *in cis* (Solnic, 1985) or *in trans* (Eul *et al.*, 1995), bringing the 18th and 24th exons into close proximity (Figure 4B). This structure could sequester the consensus splice donor and acceptor of the 18th and 24th exons of the major transcript, permitting nearby cryptic splice sites to be activated to

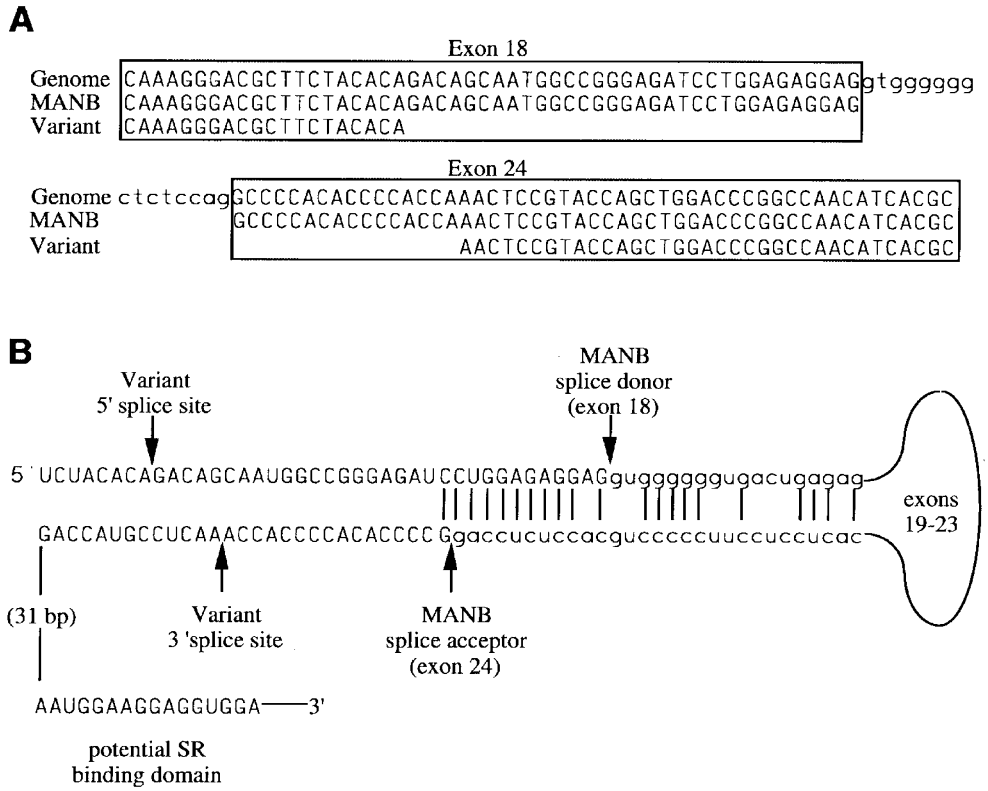

Figure 4. Partial exon sequence alignment corresponding to the deletion in the variant cDNA. A. Partial sequences flanking the junction site. Capital letters represent exon sequences and lower case letters are intron sequences of the genomic DNA (genome). B. Alignment of complementary regions that could potentially form a secondary structure in the RNA to allow splicing to occur. Capital letters represent exon sequences and low case letters intron sequences. Solid bars show the complementary bases identified by computing. A purine-rich domain in the 24th exon is shown in bold, which could function as a SR protein-binding site.

generate the 2.3 kb message.

The junctions where the 704 bp deletion occurred were not flanked by a consensus splice donor or acceptor. Since there was a CACA in the 18th exon and a CA in the 24th exon flanking the junction, the second A or CA could be from either exon (Figure 4A). Thus the nucleotides flanking the junction site could be GA-CA, AG-CC, or CA-AC. Regardless of which pair of nucleotides are actually used, none has been found previously to mediate a splicing event. The splice sites do not belong to either the nuclear pre-mRNA GU-AG (Tarn and Steitz, 1996) introns or the minor class AU-AC introns (Hall and Padgett, 1996). Thus the 2.3 kb message appeared to be generated by a previously undescribed RNA splicing event with non-consensus splice sites.

The 2.3 kb message was not visible on conventional northern blots (Nilssen *et al.*, 1997; Sun *et al.*, 1999), and was only revealed as a minor band by RT-PCR. This indicates that the splice sites generating the 704 bp deletion were very weak, which is a common characteristic of other activated cryptic splice sites (Nelson and Green, 1990). In addition, a polypurine domain present in the 24th exon (Figure 4B) might function as a potential serine-arginine (SR)-binding domain (Watakabe *et al.*, 1992), which could also be involved in activation of the cryptic splice sites. It is also possible that the missing segment of mRNA is deleted from the normal mature mRNA after the primary transcript is processed.

Alternative splicing is a powerful post-transcriptional mechanism for regulating gene expression. By utilizing different splice sites within the same precursor RNA, multiple transcripts and the protein products they encode can be generated from a single gene (Smith *et al.*, 1989). Another lysosomal enzyme gene, β-galactosidase, also expresses an enzymatically inactive variant mRNA that encodes an elastin/laminin binding protein (Hinek *et al.*, 1993). Minor transcripts have been observed from the lysosomal α-mannosidase gene in mouse and human tissues (Liao *et al.*, 1996, Beccari *et al.*, 1997). Evidence for alternative splicing to generate variants has been found in human cells. The first reported α-mannosidase cDNA (Nebes and Schmidt, 1994) has a 69 bp deletion in the middle of the coding region compared to other human lysosomal α-mannosidase cDNAs. The boundaries of the deletion conform to the GU-AG introns, indicating that the deletion was caused by skipping one exon. Another study also isolated an alternatively spliced 3.6 kb human lysosomal α-mannosidase cDNA from spleen with an 83 bp deletion and a 691 bp insertion within the coding region (Liao *et al.*,

1996). These data and the present study suggest that the human lysosomal α-mannosidase gene might encode multiple proteins by alternative splicing. As with most other alternatively spliced variants, the potential function of the 2.3 kb human lysosomal α-mannosidase variant has not been identified.

Acknowledgments

The authors thank A. Poleski and M. Parente for technical assistance; Dr. J. Fang for sequencing and oligonucleotides; and Drs N. Fraser, P. Henthorn, M. Haskins, and S. Liebhaber for helpful discussions. This work was supported by grants from the National Institute of Health (RR02512 and DK42707) and the Mrs. Cheever Porter Foundation. H.S. was supported by a fellowship from the Robert J. and Helen C. Kleberg Foundation.

References

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1994) Current Protocols in Molecular Biology. (John Wiley and Sons, Inc., New York) pp. 4.0.1-4.10.11

Beccari, T., Appolloni, M. G., Costanzi, E., Stinchi, S., Stirling, J. L., Della Fazia, M. A., Servillo, G., Viola, M. P. and Orlacchio, A. (1997) Lysosomal α-mannosidases of mouse tissues: characteristics of the isoenzymes, and cloning and expression of a full-length cDNA. *Biochem. J.* 327: 45-49

Brunak, S. and Engelbrecht, J. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J. Mol. Biol.* 220: 49-65

Daniel, P. F., Winchester, B. and Warren, C. D. (1994) Mammalian α-mannosidases-multiple forms but a common purpose? *Glycobiology* 4: 551-566

Eul, J., Graessmann, M. and Graessmann, A. (1995) Experimental evidence for RNA *trans*-splicing in mammalian cells. *EMBO J.* 14: 3226-3235

Fu, X. Y., Ge, H. and Manley, J. L. (1988) The role of the polypyrimidine stretch at the SV40 early pre-mRNA 3' splice site in alternative splicing. *EMBO J.* 7: 809-817

Green, M. R. (1991) Biochemical mechanism of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.* 7: 559-599

Hall, S. L. and Padgett, R. A. (1996) Requirement of U12 snRNA in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* 271: 1716-1718

Hinek, A., Rabinovitch, M., Keeley, F., Okamura-Oho, Y. and Callahan J. (1993) The 67-kD elastin/laminin-binding protein is related to an enzymatically inactive, alternatively spliced form of β-galactosidase. *J. Clin. Invest.* 91: 1198-1205

Kaneda, Y., Hayes, H., Uchida, T., Yoshida, M. C. and Okada, Y. (1987) Regional assignment of five genes on human chromosome. *Chromosoma* 95: 8-12

Krainer, A. R. and Maniatis, T. (1988) In Frontiers in molecular biology: transcription and splicing. B. D. Hames, D. M. Glover, eds. (IRL Press, Oxford/Washington D.C) pp.131-206

Liao, Y. F., Lal, A. and Moremen, K. W. (1996) Cloning, expression, purification, and characterization of the human broad specificity lysosomal acid α-mannosidase. *J. Biol. Chem.* 271: 28348-28358

Lowery, D. E. and Van Ness, B. G. (1988) Comparison of in vitro and in vivo splice site selection in immunoglobulin precursor mRNA. *Mol. Cell Biol.* 8: 2610-2619

Nebes, V. L. and Schmidt, M. C. (1994) Human lysosomal α-mannosidase: Isolation and nucleotide sequence of the full-length cDNA. *Biochem. Biophys. Res. Comm.* 200: 239-245

Nelson, K. K. and Green, M. R. (1990) Mechanism for cryptic splice site activation during pre-mRNA splicing. *Proc. Natl. Acad. Sci.* USA, 87: 6253-6257

Nilssen, O., Berg, T., Riise, H. M. F., Ramachandran, U., Evjen, G., Hansen, G. Malm, D., Tranebjerg, L. and Tollersrud, O. K. (1997) α-mannosidosis: Functional cloning of the lysosomal α-mannosidase cDNA and identification of a mutation in 2 affected siblings. *Hum. Mol. Genet.* 6: 717-726

Riise, H. M. F., Berg, T., Nilssen, O., Romeo, G., Tollersrud, O. K. and Ceccherin, I. (1997) Genomic structure of the human lysosomal α-mannosidase gene (MANB). *Genomics* 42: 200-207

Shapiro, M. B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15: 7155-7174

Smith, C. W., Patton, J. G. and Nadal-Ginard, B. (1989) Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* 23: 527-577

Solnic, D. (1985) Alternative splicing caused by RNA secondary structure. *Cell* 43: 667-676

Sun, H., Yang, M., Haskins, M. E., Patterson, D. F. and Wolfe, J. H. (1999) Retrovirus vector-mediated correction and cross-correction of lysosomal α-mannosidase deficiency in human and feline fibroblasts. *Hum. Gene Ther.* 10: 1311-1319

Tarn, W. Y. and Steitz, J. A. (1996) A novel spliceosome containing U11, U12, and U15 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* 84: 801-811

Thomas, G. H. and Beaudet, A. L. (1995) Disorders of glycoprotein degradation and structure: α-mannosidosis, β-mannosidosis, sialidosis,aspartylglucosaminuria, and carbohydrate-deficient glycoprotein syndrome. In Metabolic Basis of Inherited Disease, 7th Ed. C.R. Scriver, A.L. Beaudet, WS. Sly, D. Valle, eds. (McGraw-Hill, New York) pp.2529-2561

Watakabe, A., Tanaka, K. and Shimura, Y. (1992) The role of exon sequences in splice site selection. *Genes & development* 7: 407-418