

High Level Transcription of the Glucocerebrosidase Pseudogene in Normal Subjects and Patients with Gaucher Disease

J. Sorge,* E. Gross,* C. West,† and E. Beutler†

*Stratagene, La Jolla, California 92037; and †Department of Molecular and Experimental Medicine, Scripps Clinic and Research Foundation, La Jolla, California 92037

Abstract

Gaucher disease is due to mutations involving the glucocerebrosidase gene. A closely homologous pseudogene is located ~ 16 kD downstream from the functional gene. Sequence analysis of clones from cDNA libraries made from skin fibroblast cultures showed several independent clones with the sequence of an aberrantly processed pseudogene message. Examination of cellular RNA from lymphoblasts or fibroblasts obtained from thirteen Gaucher disease patients, one Gaucher disease heterozygote, and four normal subjects showed that the pseudogene was consistently transcribed, and that in some cases the level of transcription seemed to be approximately equal to that of the functional gene. The transcription of the pseudogene must be taken into account when attempting to detect mutations of glucocerebrosidase by the study of cDNA libraries. (*J. Clin. Invest.* 1990. 86:1137-1141.) Key words: polymerase chain reaction • splicing • mutation • gene

Introduction

Gaucher disease is the most common of the glycolipid storage diseases. It is caused by an inherited deficiency of the acid glycohydrolase glucocerebrosidase. Recently, methods for the molecular diagnosis of this disorder using the polymerase chain reaction (PCR) have been developed (1, 2). However, the evaluation of PCR results is confounded by the presence of a glucocerebrosidase pseudogene ~ 16 kb downstream from the functional gene (3, 4). This pseudogene is not only closely homologous to the glucocerebrosidase gene, but the sequence of a portion of the normal pseudogene is identical with one of the mutations of the true gene that causes Gaucher disease.

A potential means of circumventing this difficulty might be to examine cDNA instead of genomic DNA. In investigating this approach to the evaluation of the glucocerebrosidase gene in cultured fibroblasts and transformed lymphoblasts, we have discovered that the glucocerebrosidase pseudogene is regularly transcribed at a high level in these cells. Examination of the splicing pattern of the pseudogene mRNA has provided addi-

tional insights into the effect of mutations of the splice consensus sequences on splicing.

Methods

Sequence analysis of cDNA clones. Using standard methods (5), a cDNA library was prepared in lambda Zap I from skin fibroblast cultures obtained from two unrelated patients with Type I Gaucher disease who had been found to have the common Jewish mutation at NT 1226 on one allele and none of the known mutations on the other. We use a question mark to designate a Gaucher disease allele, as defined by phenotypic expression, but one that does not contain the mutations that have been described at nucleotides 476, 764, 1090, 1266, 1361, 1448, or 1504. Hence, these subjects are designated as possessing the 1226/? genotype. The libraries were screened with oligonucleotide probes specific to the 5' (5'-AGGATAGAGGATCCACTAAA-3') and 3' (5'-CGTCCAGTGATGGAGCAGA-3') ends of the glucocerebrosidase cDNA. The sequences are identical in the active gene and the pseudogene. Selected clones were sequenced using the chain termination method (6). Nucleotides were numbered from the upstream ATG (7, 8), the A being assigned number 1.

Studies of cellular RNA by the polymerase chain reaction (PCR). Total RNA was isolated from lymphoblasts or fibroblasts from thirteen Gaucher disease patients, one Gaucher disease heterozygote, and four normal subjects. The cells were dissolved in guanidine thiocyanate, and the RNA was pelleted through a cesium chloride cushion. After phenol-chloroform extraction and ethanol precipitation, the RNA concentration was estimated by measuring the optical density at 260 nm. Poly A⁺ RNA was isolated from total RNA by binding to and then eluting from an oligo d(T) cellulose column.

3-5 µg of total RNA or 50-200 ng of Poly A⁺ RNA were alcohol precipitated and lyophilized to serve as templates for first strand cDNA synthesis. The RNA pellets were resuspended in 8 µl diethylpyrocarbonate-treated H₂O containing 250 ng of a 3' oligonucleotide primer complementary to nucleotides 1621-1639 of the glucocerebrosidase cDNA and heated to 65°C for 5 min. Then, 12 µl of a first strand cDNA synthesis reaction mixture was added so that the 20-µl system contained 0.5× PCR buffer (9), 1 mM each of dATP, dGTP, dTTP and dCTP, 1 U of RNase block (Stratagene, La Jolla, CA) and 25 U of AMV reverse transcriptase. After incubation at 42°C for 60 min and 65°C for 5 min, 20 µl of the first strand cDNA was diluted with 80 µl of a PCR reaction mix containing 250 ng of the 5' oligonucleotide extending from nucleotide 1225 to 1244 of the glucocerebrosidase cDNA and final concentration of 0.5× PCR buffer (9), 5% DMSO, 1 mM each dNTP, and 3.5 U of Taq polymerase. These primers flank the 55 nucleotide deletion in the pseudogene. Thus, a cDNA strand from the pseudogene would create an amplified segment 360 bp in length and a 415-bp segment would be formed when the cDNA had been derived from mRNA from the true gene. Contaminating genomic DNA would give a segment 879 nucleotides in length for the true gene and 824 nucleotides for the pseudogene because two introns are present between the two primers used. PCR was carried out for 30 cycles by denaturing at 92°C, 30 s, annealing at 58°C for 30 s, and extending at 72°C for 30 s.

This is publication number 6287-MEM from the Research Institute of Scripps Clinic.

Received for publication 6 March 1990 and in revised form 30 May 1990.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/90/10/1137/05 \$2.00

Volume 86, October 1990, 1137-1141

After PCR, the fragments were phenol-chloroform extracted, ethanol precipitated, and redissolved in 100 μ l of 5T.1E. 5–10 μ l were loaded on 0.7% acrylamide gel in 0.5 \times TBE and electrophoresed at 25 mA for \sim 1 h. The gels were stained with ethidium bromide.

Results

Analysis of cDNA clones. The results of sequence analysis of cDNA clones are summarized in Table I and in Fig. 1.

One of the clones (310-20A) derived from the first patient (M.J.), was sequenced from nucleotide 521 to 2151. The sequence corresponded exactly to that of normal cDNA as published previously. A second clone (310-5) from this patient was sequenced only in exons 9 and 10 and was found to have a sequence that corresponded to normal with the exception that nucleotide 1226 was G instead of A. The 5' end of a third clone (310-20C) consisted of the last 54 bp of intron 4, exon 5, and the remaining exons (3) to nt 1673. The sequence that was found was identical to that of the exons of the pseudogene except at nt 812 where a C, like that of the normal gene, was found rather than the G found in the pseudogene.

Three independent clones from the second patient (G.S.) were also investigated. One of these GSII-13 was a full length cDNA that contained the mutation at nucleotide 1226, and was sequenced only in the area of this mutation. With the exception of this expected A \rightarrow G substitution, the sequence was normal. Two other clones extended from upstream of the first start codon to the poly A tail. Both lacked exon 2, the first 106 nucleotides of exon 3, and exon 4. Both had the 55-bp deletion in exon 9 that is characteristic of the pseudogene. One of the clones GSII-9 was sequenced in its entirety and, except for the missing portions alluded to above, had a sequence identical to the exons of the pseudogene.

Analysis of fragments produced in the polymerase chain reaction. The results of a typical experiment are shown in Fig. 2 and all of the data are summarized in Table II. The 360-bp segment, representing amplification of the pseudogene-derived cDNA was present in all samples in addition to the expected

415-bp fragment. To identify whether the amplified fragments actually represented glucocerebrosidase or glucocerebrosidase pseudogene sequences, they were transferred from the gel to nitrocellulose by the method of Southern (10). They were then probed with a 2.2-kb 32 P-labeled glucocerebrosidase probe. Both the 360- and the 415-bp fragments hybridized with the probe.

Further verification of the identity of the amplified DNA segments was achieved by eluting each fragment from the gel and reamplifying each under the same conditions that had been used in the first amplification. Each fragment was then digested with Bam HI. There is one restriction site for this enzyme at nucleotide 1629. The enzyme should cleave the amplified segment from the true gene into fragments 311 and 104 bp in length, while the pseudogene-derived fragment would be 256 and 104 bp long. The expected bands were found on electrophoresis of the digests made from the reamplified 415 and 360 bp fragments.

There was considerable variation in the relative proportions of pseudogene cDNA and functional gene cDNA produced by PCR, reflecting, presumably, the amount of mRNA in the original samples. In some instances, the amount of pseudogene and functional gene DNA was approximately equal; in other samples, the amount of DNA produced from the functional gene cDNA was approximately one order of magnitude larger than the amount of DNA from the pseudogene. These findings were reproducible both when the same RNA sample served as a starting point and when another RNA preparation was made. It did not appear to depend on whether fibroblasts or lymphoblasts served as the source material.

Discussion

Pseudogenes may be defined as homologues of normal genes that no longer make a gene product. Pseudogenes with introns presumably arise as a result of tandem duplication of genes with a loss of function as a result of divergent evolution; the

Table I. Composition of the cDNA Clones from Two Subjects with Gaucher Disease

Patient (Genotype)	Clone	Nucleotides sequenced	Result	Interpretation
M.J. (1226/?)	310-20A	521–2151	Normal cDNA sequence, partial clone.	Normal gene transcript
M.J. (1226/?)	310-5	1105–1550	Normal cDNA sequence except for 1226 A \rightarrow G, partial clone.	1226 mutant gene
M.J. (1226/?)	310-20C	2985(g)–1673	Begins with last 54-bp intron 4. Pseudogene sequence except at nt 812. Poly A tail after nt 1673.	Pseudogene transcript
G.S. (1226/?)	GSII-13	1105–1550	Normal cDNA sequence except for 1226 A \rightarrow G full-length clone.	1226 mutant gene transcript
G.S. (1226/?)	GSII-9	–177–2151	Pseudogene sequence. Missing all of exon 2, first 106 nucleotides of exon 3, and all of exon 4. 55-bp deletion from exon 9. Extends to nt 2166 then poly A tail.	Misspliced pseudogene transcript

In each case, one allele contained the common mutation of nt 1226, and the mutation on the other allele could not be identified. g, genomic sequence.

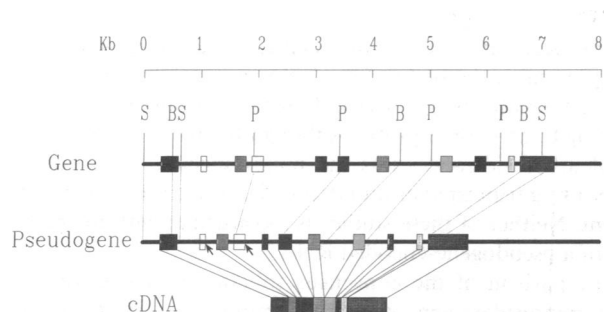


Figure 1. The structure of pseudogene-derived cDNA clone GSII-9. The functional gene and pseudogene are shown with homologous restriction sites connected by dotted lines. *S*, Sac I; *P*, Pst I; *B*, Bam HI. The arrows point to the destroyed consensus splice donor sites. Exons 2 and 4, shown as open boxes are not represented in the cDNA. They have been spliced from the messenger during processing. A portion of exon 3 is also missing. All other homologues of the exons are correctly represented in the spliced cDNA but the sequence is exactly that of the pseudogene.

existence of the normal copy of the gene may suffice for the needs of the organism, so that mutations of one of the copies is tolerated. "Processed" pseudogenes lack introns, and presum-

ably are written into the genome from processed mRNA by reverse transcriptase.

The glucocerebrosidase pseudogene contains introns and presumably arose from the glucocerebrosidase gene by tandem duplication. This event apparently occurred relatively recently in evolution. The mouse genome does not appear to contain the pseudogene (11), and the human pseudogene has a high degree of homology with the true gene, particularly so at the 3' end.

Most pseudogenes are presumably transcriptionally and translationally silent, but there are exceptions. Transcripts of the interferon pseudogene have been found (12), and the processed phosphoglycerate kinase pseudogene is transcribed and translated into enzyme in the testis (13). The latter might therefore be designated a pseudo-pseudogene.

When tandem duplication of a gene occurs as the first step in the formation of a pseudogene the duplicated genes are both functional. When the gene that is fated to become the pseudogene loses its function, there is no longer any selective pressure against further changes. For this reason, its promoter regions would eventually be expected to lose the ability to initiate transcription. The glucocerebrosidase pseudogene is certainly nonfunctional in the production of enzyme. Numerous point

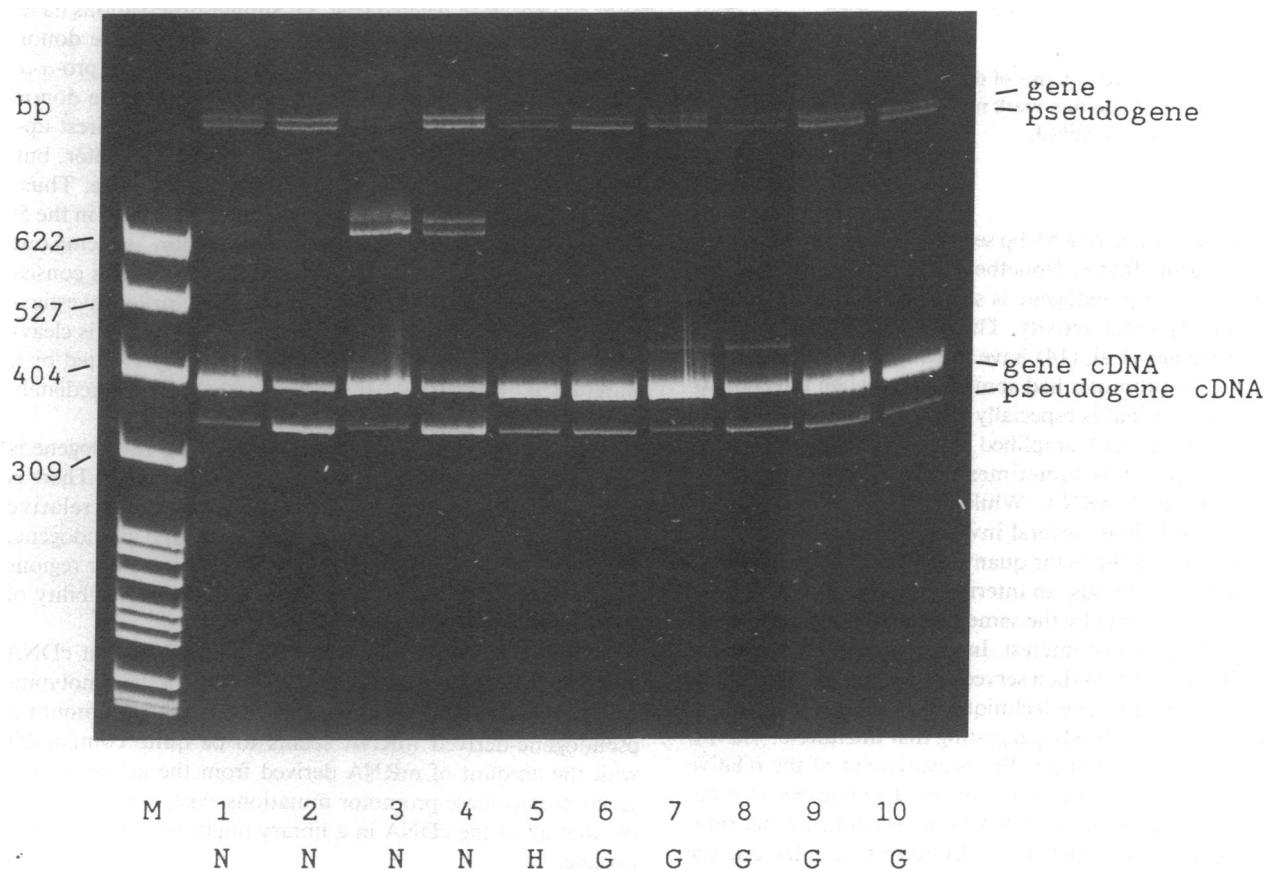


Figure 2. Acrylamide gel electrophoresis of pBR322/Msp I marker (lane *M*) and the products of PCR. The first strand cDNA template used for PCR in the channels marked *N* was from subjects without Gaucher disease. Channel 5 (*H*) template was from a heterozygote for Gaucher disease with the $?/N$ genotype. The channels marked *G* used as template first strand cDNA from patients with Gaucher disease with the following genotypes: (6) 1226/1226; (7) 1226/?; (8) 1226/?; (9) 1226/?; (10) 1226/?. The bands sized 415 and 360 bp are formed from mRNA produced by the true gene and by the pseudogene, respectively. The size difference is due to the 55-bp deletion known to exist in exon 9 of the pseudogene. The much larger fragments represent 879- and 824-bp fragments DNA amplified from genomic gene and pseudogene DNA, respectively, contaminating the preparation, since there are two introns between the two primers. The doublets seen faintly in channels 1 and 7 and distinctly in channels 3 and 4 are of unknown origin.

Table II. The Results of Performing the PCR on RNA from Fibroblasts or Lymphoblasts Cell Cultures from Gaucher Disease Patients, a Heterozygote, and Normals

Patient	Gaucher genotype	Gaucher phenotype	PCR band intensity	
			415 bp	360 bp
A.B.*	1226/?	Type I	+++	+
M.J.*	1226/?	Type I	+++	+
A.H.	1226/?	Type I	++	+
D.F.	1226/?	Type I	+++	+
A.M.	1226/?	Type I	++	+
M.H.	1226/?	Type I	+++	+
T.M.	1226/?	Type I	++	+
S.W.	1226/?	Type I	+++	+
M.K.	1226/1226	Type I	+++	+
E.C.	?/?	Type I	+++	+
K.C.	1448/?	Type I	+++	+
M.M.	1448/1448	Type II	+++	+
1260	1448/1361	Type II	+++	+
P.E.	+/?	Heterozygote	+++	+
J.P.	+/+	Normal	+	+
W.C.	+/+	Normal	+	+
T.B.	+/+	Normal	+++	+
D.G.	+/+	Normal	+++	+

The genotypes are defined in terms of the nucleotide at which the substitution occurred. Question mark indicates that none of the known mutations were identified.

* Siblings.

mutations and the loss of a 55-bp segment have deprived it of a long open reading frame. Nonetheless, it is apparent that the glucocerebrosidase pseudogene is so young that it has not yet lost all transcriptional activity. This may have been predictable, since Reiner et al. (14) have already demonstrated that the pseudogene promoter had some activity when attached to a reporter gene. What is especially remarkable is that judging by the amount of DNA amplified using PCR, the amount of pseudogene mRNA is sometimes quite comparable to the amount of normal mRNA. While PCR is not ordinarily a quantitative technique, several investigators (15–17) have recently proposed methods for quantitation of mRNA by PCR. In each of these methods, an internal standard is used. It consists of RNA amplified by the same primers that are employed to amplify the RNA of interest. In each case, the ratio of the signals of the two RNAs then served as a means of quantitating the RNA of interest. The technique used here is identical in principle to these methods, suggesting that the ratio of the two signals to one another is a valid measurement of the relative amounts of the two types of messenger. The number of pseudogene clones we found in cDNA libraries confirms that transcription occurs at a high level. Moreover, it rules out the possibility that the pseudogene-derived messenger was actually a result of allele skipping.

Two groups of investigators have reported the analysis of glucocerebrosidase poly A⁺ RNA using S1 nuclease and RNase I analysis. Graves et al. (18) demonstrated the existence of a poly A⁺ RNA that is truncated at the 3' end. However, the nondenaturing gels used in the analysis would not detect in-

sertions, deletions, alternative splicing, or nucleotide substitutions present in the body of the RNA molecule. Reiner et al. (14) used denaturing gel electrophoresis to examine RNAs hybridizing to an exon 3 probe. Although the band representing complete hybridization was the major species, other bands representing truncated or mutated variants of exon 3 were also present. Neither of these studies is inconsistent with the presence of a pseudogene-encoded RNA.

Comparison of the consensus splicing sequences of the glucocerebrosidase gene with the homologous portions of the pseudogene reveals that the 5' (donor) sites of introns 2 and 4 have been destroyed (3) (Fig. 1, arrows). In each case, the GpT that must begin each intron has been changed to a GpA. In none of the other introns have the 6 nucleotides that form the 5' consensus (19) been altered. In the case of the 3' (acceptor) sites, none of the critical ApG sequences have been altered and only a single alteration is present in any of the 15 nucleotides that form the consensus sequence (19). This changes a C → T at the nucleotide -14 of the acceptor site of intron 7, does not change the strength of the consensus; T is the preferred nucleotide at this site (19). The fact that exon 2, the 5' portion of exon 3, and exon 4 were missing from processed transcripts provides some interesting insights into the mechanism of splicing. In each case, the exon upstream of the destroyed donor site has been deleted (Fig. 1). Similar observations have recently been reported in the cases of mutated splice donor sites of the β subunit of leukocyte integrins (20) and of pro- α - α collagen (21). In each case, the exon upstream of the donor splice site mutation was deleted. Apparently, the nearest upstream donor site does not use its own normal acceptor, but rather one downstream from the mutated donor site. Thus, processing behaves as if so-called "receptor" sites scan in the 5' direction until they encounter a normal "donor," splicing out exons 5' to mutated donor sites. This phenomenon is consistent with the findings of Aebi et al. (22) and earlier investigators. Apparently, the first step in the splicing process is cleavage of the 5' splice site. When this cleavage is prevented by a mutation, a "super lariat" that contains the exon immediately upstream from the 5' mutation is formed.

The transcription of the glucocerebrosidase pseudogene is probably quite unusual although not entirely unique. There is considerable subject-to-subject variation in the relative amounts of transcripts derived from gene and pseudogene. This could be due to polymorphisms in the promoter regions of the pseudogene or to polymorphisms affecting stability of pseudogene-derived mRNA.

The practical consequence of our findings is that cDNA libraries must be used with extreme caution for the genotyping of patients with Gaucher disease; in some cases, the amount of pseudogene-derived mRNA seems to be quite comparable with the amount of mRNA derived from the active gene. If glucocerebrosidase promoter mutations exist, it is even possible that all of the cDNA in a library might be from the pseudogene.

Acknowledgments

This work was supported by the Sam Stein and Rose Stein Charitable Trust Fund, March of Dimes grant 1-955, and National Institutes of Health grants DK36639 and RR00833.

References

1. Zimran, A., J. Sorge, E. Gross, M. Kubitz, C. West, and E. Beutler. 1989. Prediction of severity of Gaucher's disease by identification of mutations at DNA level. *Lancet*. 2:349-352.
2. Theophilus, B., T. Latham, G. A. Grabowski, and F. I. Smith. 1989. Gaucher disease: molecular heterogeneity and phenotype-genotype correlations. *Am. J. Hum. Genet.* 45:212-225.
3. Horowitz, M., S. Wilder, Z. Horowitz, O. Reiner, T. Gelbart, and E. Beutler. 1989. The human glucocerebrosidase gene and pseudogene: structure and evolution. *Genomics*. 4:87-96.
4. Zimran, A., J. Sorge, E. Gross, M. Kubitz, C. West, and E. Beutler. 1990. A glucocerebrosidase fusion gene in Gaucher disease. Implications for the molecular anatomy, pathogenesis, and diagnosis of this disorder. *J. Clin. Invest.* 85:219-222.
5. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *In* Molecular cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
6. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463-5467.
7. Sorge, J., C. West, B. Westwood, and E. Beutler. 1985. Molecular cloning and nucleotide sequence of the human glucocerebrosidase gene. *Proc. Natl. Acad. Sci. USA*. 82:7289-7293.
8. Sorge, J. A., C. West, W. Kuhl, L. Treger, and E. Beutler. 1987. The human glucocerebrosidase gene has two functional ATG initiator codons. *Am. J. Hum. Genet.* 41:1016-1024.
9. Kogan, S. C., M. Doherty, and J. Gitschier. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences: application to hemophilia A. *N. Engl. J. Med.* 317:985-990.
10. Southern, E. 1979. Gel electrophoresis of restriction fragments. *Methods Enzymol.* 68:152-176.
11. O'Neill, R. R., T. Tokoro, C. A. Kozak, and R. O. Brady. 1989. Comparison of the chromosomal localization of murine and human glucocerebrosidase genes and of the deduced amino acid sequences. *Proc. Natl. Acad. Sci. USA*. 86:5049-5053.
12. Goeddel, D. V., D. W. Leung, T. J. Dull, M. Gross, R. M. Lawn, R. McCandliss, P. H. Seeburg, A. Ulrich, E. Yelverton, and P. W. Gray. 1981. The structure of eight distinct cloned human leukocyte interferon cDNAs. *Nature (Lond.)*. 290:20-26.
13. McCarrey, J. R., and K. Thomas. 1987. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature (Lond.)*. 326:501-505.
14. Reiner, O., and M. Horowitz. 1988. Differential expression of the human glucocerebrosidase-coding gene. *Gene (Amst.)*. 73:469-478.
15. Becker-André, M., and K. Hahlbrock. 1989. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res.* 17:9437-9446.
16. Wang, A. M., M. V. Doyle, and D. F. Mark. 1989. Quantitation of mRNA by the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA*. 86:9717-9721.
17. Gilliland, G., S. Perrin, K. Blanchard, and H. F. Bunn. 1990. Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proc. Natl. Acad. Sci. USA*. 87:2725-2729.
18. Graves, P. N., G. A. Grabowski, M. D. Ludman, P. Palese, and F. I. Smith. 1986. Human acid beta-glucosidase: Northern blot and S1 nuclease analysis of mRNA from HeLa cells and normal and Gaucher disease fibroblasts. *Am. J. Hum. Genet.* 39:763-774.
19. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459-472.
20. Kishimoto, T. K., K. O'Connor, and T. A. Springer. 1989. Leukocyte adhesion deficiency. Aberrant splicing of a conserved intron sequence causes a moderate deficiency phenotype. *J. Biol. Chem.* 264:3588-3595.
21. Weil, D., M. Bernard, N. Combates, M. K. Wirtz, D. W. Hollister, B. Steinmann, and F. Ramirez. 1988. Identification of a mutation that causes exon skipping during collagen pre-mRNA splicing in an Ehlers-Danlos syndrome variant. *J. Biol. Chem.* 263:8561-8564.
22. Aebi, M., H. Hornig, R. A. Padgett, J. Reiser, and C. Weissmann. 1986. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell*. 47:555-565.