A Glucocerebrosidase Fusion Gene in Gaucher Disease

Implications for the Molecular Anatomy, Pathogenesis, and Diagnosis of this Disorder

A. Zimran,* J. Sorge,‡ E. Gross,‡ M. Kubitz,‡ C. West,* and E. Beutler*

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

Abstract

The molecular diagnosis of Gaucher disease has been difficult due to the existence of several different point mutations in the glucocerebrosidase gene and due to the presence of a tightly linked, highly homologous pseudogene. We now report the occurrence of a "Lepore-like" glucocerebrosidase fusion gene in which the 5' end is the functional gene and the 3' end is the pseudogene. This further complicates the molecular diagnosis of Gaucher disease but sheds light on the molecular anatomy of the glucocerebrosidase gene complex and on the pathogenesis of this important storage disease. (*J. Clin. Invest.* 1990. 85:219–222.) crossover • Lepore • mutation • pseudogene

Introduction

Gaucher disease is probably the most common glycolipid storage disease. It is caused by an inherited deficiency of the lysosomal enzyme glucocerebrosidase (1). In addition to the functional gene, a glucocerebrosidase pseudogene has been identified; both the functional gene and the pseudogene appear to be on chromosome 1 (2, 3) and both have been sequenced in their entirety (4).

Four different single-base substitutions have been reported to cause this disease (Table I) (5–8). In some patients disease-producing alleles remain unidentified, however. We therefore cloned and sequenced cDNA from cultured skin fibroblasts or lymphoblasts from such individuals. In the course of these studies, we have sequenced a cDNA representing transcript from a fusion gene in which the 5' end is the functional glucocerebrosidase gene and the 3' end is the pseudogene.

In this study we characterize this fusion gene and discuss its importance for understanding the molecular anatomy of the glucocerebrosidase gene complex and its implications for the molecular pathogenesis and diagnosis of Gaucher disease.

Methods

Patient studied. The patient (D.W.) is a 21-yr-old female with type 1 (adult type) Gaucher disease. The diagnosis was made at age 3 yr, based on the clinical findings of hepatosplenomegaly, on the identification of

---

Gaucher cells in bone marrow aspiration, and on the demonstration of a low activity of the acid β-glucosidase in her white blood cells.

Splenectomy was performed at age 9 yr, after deterioration of her blood counts with frequent bleeding manifestations. Episodic bone pains have occurred since age 10, however, she has not suffered fractures. At present the most prominent manifestation of her Gaucher disease is massive hepatomegaly with mild abnormalities of her liver function tests.

Cells and RNA. Skin fibroblasts were obtained from a skin biopsy and grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Cellular RNA was prepared using the guanidium thiocyanate/cesium chloride method (9); and the poly(A) rich fraction of RNA was prepared by two purifications with oligo (dT) Sepharose (10).

Preparation and screening of cDNA library. A cDNA library was prepared and screened as previously described (11) with the following modifications. We ligated the cDNA to λZAP (12) and screened our library with two 20-mer oligonucleotide probes (one made to the 5' end of the cDNA and the other to the 3' end so that we could select full-length clones).

The full length clones were sequenced on both strands using modifications of the chain termination dideoxy sequencing technique (13) with a sequenase kit (International Biotechnology, Inc., New Haven, CT) or an AMV reverse transcriptase sequencing kit (Stratagene, La Jolla, CA) according to the recommendations of the suppliers. A set of 17–20-mer oligonucleotides were used as primers.

Analysis of genomic DNA. High molecular weight genomic DNA was prepared from white blood cells according to established methods (14).

The polymerase chain reactions were performed as described (15) for identification of the crossover site of the fusion gene.

Ssp I digests were made by incubating 10 μg of DNA in 400 μl of a buffer (10 mM Tris-Cl pH 7.5, 105 mM NaCl 10 mM Mg Cl$_2$, 1 mM DTT, 0.02% gelatin). Another 120 U of Ssp I was then added and incubation was continued for an additional 2 h. Electrophoresis was performed in 0.36% ultrapure agarose (Bethesda Research Laboratories, Gaithersburg, MD) poured on a 3-mm 1% base in 40 mM Tris-Cl, 20 mM Na acetate, 18 mM NaCl, 2 mM EDTA, pH 8.05 buffer. Electrophoresis was carried out for 18 h at 2 V/cm. The gel was then treated with 0.25 M HCl for 15 min alkalinized with 0.5 M NaOH in 1.5 M NaCl for 1 h and finally treated with 0.5 M Tris pH 7.5 in 3 M NaCl for 2 h. The DNA was transferred from the gel to nitrocellulose for 3 d in 10× SSC (14). The gels were developed with the cDNA probe described previously (16), applying 35,000 cpm/mm^2 of nitrocellulose in 12 ml of annealing mix (50% formamide, 1× Denhardt's solution, 10 mM Hepes, pH 7.4, 3× SSC, 0.1 mg/ml salmon sperm DNA [denatured]) at 42°C. The filter was washed at 50–55°C in 350 ml of 0.1× SSC, 0.1% SDS three times for 1 h each.

Sac II digests were performed by adding 200 U of Sac II to 500 μl of buffer (6 mM Tris-Cl pH 7.5, 6 mM MgCl$_2$, 6 mM NaCl, 1 mM DTT, 0.02% gelatin) containing 10 μg of DNA, incubating for 4 h at 37°C. The gel was treated with acid, neutralized, and transferred to nitrocellulose in a manner identical to that used for the Ssp I digest. The filter was developed using the same probe at 55,000 cpm/mm^2.

Results

Of the 18 positive cDNA clones that were obtained 5 contained the entire glucocerebrosidase coding regions. We se-

Table I. Mutations in Gaucher Disease

Position	Base substitution normal → mutant	Amino acid change normal → mutant	Restriction site created	Disease type	Reference
476	G → A	Arg → Gln	Bst NI, Eco RII	I	(8)
1226	A → G	Asn → Ser	—	I	(6)
1361	C → G	Pro → Arg	Hha I	II	(7)
1448	T → C	Leu → Pro	Nci I	I, II, III	(5)

quenced two different clones, representing each of the two alleles.

While one of the clones had a single base substitution (A → G) at position 1226, the second one was found to have three different base pair substitutions in the coding region: T → C at position 1448 (Leu → Pro), G → C at position 1483 (Ala → Pro), and G → C at position 1497 (no amino acid change). A fourth nucleotide substitution G → A occurred at position 1703 in the 3' noncoding region.

All of these four point mutations match exactly the sequence found normally in the glucocerebrosidase pseudogene downstream from nucleotide 1300 on the cDNA sequence (4, 11). To study the molecular mechanism of the apparent recombination of active gene and pseudogene we performed Southern blot analysis of genomic DNA, digested with the restriction endonucleases Sac II and Ssp I separately (Fig. 1). An attempt to improve the resolution of the pattern on the gel by using rotating agarose gel electrophoresis (17) and reverse field electrophoresis using the PPI-200 programmable electrophoresis controller (MJ Research, Inc., Cambridge, MA) under various conditions failed to improve resolution of fragments in the desired size range. Sac II normally cuts outside

the glucocerebrosidase gene/pseudogene complex, giving a single band of ~ 49 kilobases (kb) when hybridizing to a full-length glucocerebrosidase cDNA probe. In the present case, a second shorter fragment of ~ 27 kb was obtained. Similar Southern blot analysis with Ssp I restriction endonuclease normally reveals two bands of ~ 18 and 13 kb. In the present case, an additional band of 14 kb was obtained.

These studies provide evidence for rearrangement of DNA in the glucocerebrosidase gene complex. Unequal crossover apparently has occurred between the functional glucocerebrosidase gene and the pseudogene as the molecular event responsible for the creation of the fusion gene. The loss of DNA between the gene and pseudogene accounts for the shorter than normal fragments found on restriction analyses.

In determining the crossover region, we analyzed genomic DNA by the polymerase chain reaction technique, using our previously described methods for the detection of mutations 1226 and 1448 (15). This analysis indicates that the transition from the functional gene to the pseudogene occurred between nucleotides 5933 and 6272 of the glucocerebrosidase gene. Since the sequences of the gene and pseudogene are identical between these two nucleotides, the exact point of crossover could not be determined.

Discussion

We report the occurrence in a patient with adult-type Gaucher disease of a cDNA representing transcription from a fusion gene, in which the 5' end is the functional glucocerebrosidase gene and the 3' end is the pseudogene.

A map of the glucocerebrosidase gene/pseudogene complex reveals that the restriction endonuclease Sac II cuts outside the complex, yielding a single fragment ~ 49 kb in length (Fig. 2). If the fusion gene had arisen through the process of gene conversion (18), the length of this fragment should remain unaltered in the affected individuals. Similarly, if a long transcript encompassing both the gene and the pseudogene had been processed in such a way as to create the mRNA that was detected, the restriction pattern would have been unaltered. The appearance of an altered, shorter fragment in the Sac II digest of our patient proves that the fusion gene was created through an unequal crossover event.

This conclusion is further confirmed by the results of the Ssp I analysis. Normally (in healthy individuals and in Gaucher patients), Ssp I cuts outside the glucocerebrosidase gene complex and inside the intergenic region, producing two major bands when hybridizing to the glucocerebrosidase cDNA. Since the intergenic sequence is missing in the fusion gene described herein, there should be a loss of the Ssp I intergenic site(s) of this gene complex with a probable additional
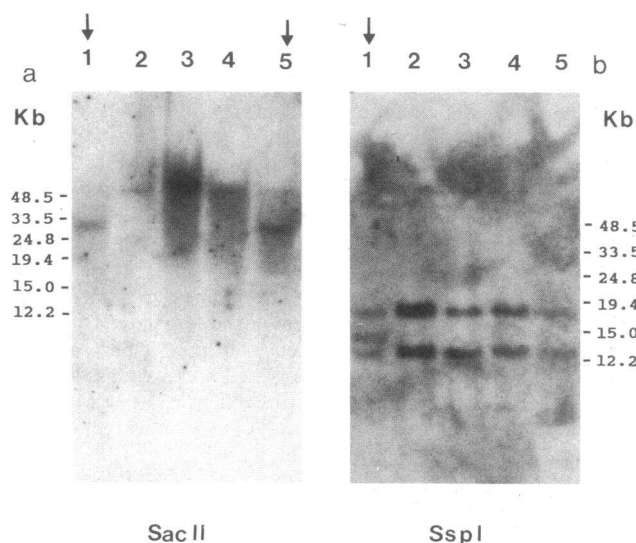

Figure 1. Southern blot analysis of genomic DNA samples digested with Sac II and Ssp I. (a) the samples from the patient are in lanes 1 and 5. Lanes 2-4, showing the normal pattern, were obtained from a normal individual (lane 2) and from two other Gaucher patients (lanes 3 and 4). (b) lane 1, DNA from the patient showing the extra 14-kb band; lanes 2, 4, and 5 contain DNA from other Gaucher disease patients; and lane 3 is a normal control. The lanes from the patient are marked with an arrow.

change in the distance between the two outside Ssp I sites. Our finding of a third band ($\sim$ 14 kb in length) fits this prediction well, and indicate that the other normally sized bands are derived from the other allele.

Ssp I analysis is technically easier to perform than Sac II analysis, since it requires blotting of shorter DNA fragments. Therefore, we used it as a means for the detection of similar fusion genes in other patients. 18 additional individuals with Gaucher disease who did not have the 1226/1226 genotype were studied. No additional cases were encountered.

A fusion gene containing the 5' portion of the δ-globin gene and the 3' part of the β-globin gene produces a form of thalassemia in which the abnormal hemoglobin is designated hemoglobin Lepore. The fusion gene we describe here is the only other inherited disease of man of which we are aware in which a fusion gene producing an abnormal protein has been identified as the result of crossing over between two different genes.

Similar events, presumably representing crossing over within a gene that generates a duplication, have been described in the case of Duchenne muscular dystrophy (19) and in the LDL receptor gene (20). Color blindness is apparently frequently due to crossing over between genes for green and red pigment (21, 22). Crossovers between virtually identical genes, the α-globin genes, accounts for the deletions seen in α-thalassemia-2 (23) but here the fusion protein is normal and the disease state results from the loss of a copy rather than the formation of an abnormal one.

The existence of this "Lepore"-like glucocerebrosidase gene is an experiment of nature that provides data regarding the molecular anatomy of the glucocerebrosidase genes on chromosome 1 and also has significant implications in respect to the molecular diagnosis and pathogenesis of Gaucher disease. The fact that the 5' end of the fusion gene was derived from the functional gene indicates that the functional gene is 5' to the pseudogene; the complex that would be formed if the

functional gene were 3' to the pseudogene would have included a copy of the normal gene, and in this event the patient would not have had Gaucher disease. Moreover, since the deletion should be equal to the size of the deleted portions of the gene itself plus the length of the intergenic DNA, we can now calculate that the functional gene is separated from the pseudogene by a distance of $\sim$ 16 kb (see Fig. 2).

The fusion gene that we have detected would drastically impair the production or the stability of glucocerebrosidase. It contains a base change occurring normally in the pseudogene that causes severe Gaucher disease when present as a point mutation in the functional gene (5). In addition another downstream base pair substitution produces another ala $\rightarrow$ pro substitution at nucleotide 1483.

The analysis of mutations in Gaucher disease patients has been difficult due to the existence of several different abnormal alleles and because of the need to separate the glucocerebrosidase gene from its 95% homologous pseudogene. The occurrence of fusion genes makes such analysis even more complicated, but fortunately crossovers of the type we have described here are uncommon. In any case, this experiment does yield interesting information regarding the structure of this important gene complex.

Acknowledgments

References

1. Brady, R. O., J. N. Kanfer, R. M. Bradley, and D. Shapiro. 1966. Demonstration of a deficiency of glucocerebroside-cleaving enzyme in Gaucher's disease. *J. Clin. Invest.* 45:1112–1115.

2. Ginns, E. I., P. V. Choudary, S. Tsuji, B. Martin, B. Stubblefield, J. Sawyer, J. Hozier, and J. A. Barranger. 1985. Gene mapping and leader polypeptide sequence of human glucocerebrosidase: Implications for Gaucher disease. *Proc. Natl. Acad. Sci. USA.* 82:7101–7105.

3. Shafit-Zagardo, B., E. A. Devine, M. Smith, G. A. F. Arredondo, and R. J. Desnick. 1981. Assignment of the gene for acid beta-glucosidase to human chromosome 1. *Am. J. Hum. Genet.* 33:564–575.

4. 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.

5. Tsuji, S., P. V. Choudary, B. M. Martin, B. K. Stubblefield, J. A. Mayor, J. A. Barranger, and E. I. Ginns. 1987. A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. *N. Engl. J. Med.* 316:570–621.

6. Tsuji, S., B. M. Martin, J. A. Barranger, B. K. Stubblefield, M. E. LaMarca, and E. I. Ginns. 1988. Genetic heterogeneity in type 1 Gaucher disease. Multiple genotypes in Ashkenazic and non-Ashkenazic individuals. *Proc. Natl. Acad. Sci. USA.* 85:2349–2352, 5708.

7. Reiner, O., M. Wigderson, and M. Horowitz. 1988. Structural analysis of the human glucocerebrosidase genes. *DNA.* 7:107–116.

8. Graves, P. N., G. A. Grabowski, R. Eisner, P. Palese, and F. I. Smith. 1988. Gaucher disease type 1: cloning and characterization of a cDNA encoding acid β-glucosidase from an Ashkenazi Jewish patient. *DNA.* 7:521–528.

9. Chirgwin, J. M., A. E. Pryzbyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active RNA from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5298.
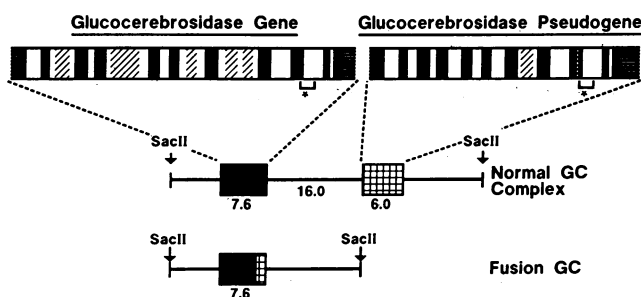

Figure 2. The formation of a fusion gene from the glucocerebrosidase (GC) gene and pseudogene. The portion of the genes in which the crossing-over occurred are shown.* Sac II restriction analysis of the glucocerebrosidase gene/pseudogene complex shows that the fragment containing the normal glucocerebrosidase gene/pseudogene complex is $\sim$ 49 kb long, and the one containing the fusion gene is $\sim$ 27 kb long. The fusion gene is of the same length as the functional GC gene (7.6 kb), because all of the missing segments of the pseudogenes (which make it $\sim$ 1.6 kb shorter) are 5' to the point of crossing-over. Most of these segments are Alu sequences (4) and are represented as diagonal dashed areas on the GC gene diagram. The 55 bp deletion in exon 9 of the pseudogene is represented by white dots. The horizontal lines in exons 1 and 11 are the noncoding regions. Since the lengths of both the GC gene and pseudogene are known, as are the lengths of the Sac II fragments, the intergenic distance can be calculated as being $\sim$ 16 kb.

10. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

11. 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.

12. Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse. 1988. Lambda ZAP: A bacteriophase lambda expression vector with *in vivo* excision properties. *Nucleic Acids Res.* 16:7583–7600.

13. Sanger, F. 1981. Determination of nucleotide sequences in DNA. *Science (Wash. DC).* 214:1205–1210.

14. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *In* Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

15. Zimran, A., J. Sorge, E. Gross, M. Kubitz, C. West, and E. Beutler. 1989. Prediction of the clinical course of Gaucher disease by identification of mutations at the DNA level. *Lancet.* ii:349–352.

16. Sorge, J., T. Gelbart, C. West, B. Westwood, and E. Beutler. 1985. Heterogeneity in type I Gaucher disease demonstrated by restriction mapping of the gene. *Proc. Natl. Acad. Sci. USA.* 82:5442–5445.

17. Southern, E. M., R. Anand, W. R. A. Brown, and D. S. Fletcher. 1987. A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res.* 15:5925–5943.

18. Baltimore, D. 1981. Gene conversion: some implications for immunoglobulin genes. *Cell.* 24:592–594.

19. Hu, X., A. H. M. Burghes, D. E. Bulman, P. N. Ray, and R. G. Worton. 1989. Evidence for mutation by unequal sister chromatid exchange in the Duchenne muscular dystrophy gene. *Am. J. Hum. Genet.* 44:855–863.

20. Lehrman, M. A., J. L. Goldstein, D. W. Russell, and M. S. Brown. 1987. Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. *Cell.* 48:827–835.

21. Drummond-Borg, M., S. Deeb, and A. G. Motulsky. 1988. Molecular basis of abnormal red-green color vision: a family with three types of color vision defects. *Am. J. Hum. Genet.* 43:675–683.

22. Nathans, J., T. P. Piantanida, R. L. Eddy, T. B. Shows, and D. S. Hogness. 1986. Molecular genetics of inherited variation in human color vision. *Science (Wash. DC).* 232:203–210.

23. Embury, S. H., J. A. Miller, A. M. Dozy, Y. W. Kan, V. Chan, and D. Todd. 1980. Two different molecular organizations account for the single alpha-globin gene of the alpha-thalassemia-2 genotype. *J. Clin. Invest.* 66:1319–1325.