

## RESEARCH ARTICLE

## Molecular Analysis and Clinical Findings in the Spanish Gaucher Disease Population: Putative Haplotype of the N370S Ancestral Chromosome

Bru Cormand,<sup>1</sup> Daniel Grinberg,<sup>1</sup> Laura Gort,<sup>2</sup> Amparo Chabás,<sup>2</sup> and Lluïsa Vilageliu<sup>1\*</sup><sup>1</sup>Departament de Genètica, Facultat de Biologia, Universitat de Barcelona<sup>2</sup>Institut de Bioquímica Clínica, Corporació Sanitària Clínic, Barcelona, Spain

Communicated by Albert de la Chapelle

Gaucher disease (GD) is an autosomal recessive disorder caused by mutations in the lysosomal  $\beta$ -glucocerebrosidase (*GBA*) gene. As the disease is particularly prevalent among Ashkenazi Jews, most studies have been carried out on this ethnic group. In the current study, we present a mutation analysis of the *GBA* gene in Spanish patients together with the clinical findings. We conducted a systematic analysis in 53 unrelated GD patients. The *GBA* gene was initially scanned for nine previously described mutations by ASO hybridization or restriction analysis after PCR amplification. The remaining unidentified alleles were screened by nonisotopic PCR-SSCP analysis and sequenced. This approach allowed the identification of 101 of 106 GD alleles (95.3%) involving 24 different mutations, 11 of which are described for the first time: G113E (455G→A), T134P (517A→C), G389E (1283G→A), P391L (1289C→T), N392I (1292A→T), Y412H (1351T→G), W(-4)X (108G→A), Q169X (662C→T), R257X (886C→T), 500insT, and IVS5+1G→T. Most mutations are present in one or few GD chromosomes. However, two mutations, N370S (1226A→G) and L444P (1448T→C), are very frequent and account for 66.1% of the total number of alleles. Linkage disequilibrium was detected between these two mutations and an intragenic polymorphism, indicating that expansion of founder alleles occurred in both cases. Analysis of several microsatellite markers close to the *GBA* gene allowed us to establish the putative haplotype of the ancestral N370S chromosome. Hum Mutat 11:295–305, 1998. © 1998 Wiley-Liss, Inc.

KEY WORDS: Gaucher disease; *GBA* gene; mutation screening; linkage disequilibrium; Spanish patients; N370S mutation; ancestral haplotype

## INTRODUCTION

Gaucher disease (GD) is a lysosomal glycolipid storage disorder characterized by the accumulation of glucosylceramide in phagocytes due to deficient activity of  $\beta$ -glucocerebrosidase (E.C. 3.2.1.45). It is inherited as an autosomal recessive trait. Three major types of Gaucher disease have been described on the basis of the absence (type I) or presence and severity (types II and III) of primary central nervous system involvement. The disease is particularly frequent in the Ashkenazi Jewish population in which the incidence of GD is estimated as 1 in 850 (Beutler and Grabowski, 1995), while in other populations the frequency ranges between 1 in 40,000 and 1 in 60,000 (Grabowski, 1993).

In most of the cases described so far, GD is caused by mutations in the *GBA* gene encoding  $\beta$ -glucocerebrosidase. More than 70 mutations have

been described in the *GBA* gene to date (Beutler and Gelbart, 1997). Mutation N370S (1226A→G) is the most prevalent among Ashkenazi Jews (approximately 70% of GD alleles) (Beutler et al., 1992a; Horowitz et al., 1993; Sibille et al., 1993) and is also frequent in other populations (22–54%) (Walley et al., 1993; Cormand et al., 1995; Michelakakis et al., 1995; Amaral et al., 1996; Tylki-Szymanska et al., 1996; le Coutre et al., 1997).

Twelve polymorphic sites have been detected

Received 27 June 1997; accepted 30 September 1997.

\*Correspondence to: Lluïsa Vilageliu, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, E-08071 Barcelona, Spain; Fax: 34-3-411-0969

Contract grant sponsor: CICYT; Contract grant number: SAF 93-0479-C02-01

Bru Cormand is currently at the Folkhälsan Institute of Genetics, Helsinki, Finland.

within the *GBA* gene (Beutler et al., 1992b). All are in linkage disequilibrium and constitute only two major haplotypes. N370S has always been found associated with the – haplotype (Zimran et al., 1990a), suggesting a single origin for this mutation. The *GBA* gene was recently mapped to a 3.2-cM interval between markers D1S305 and D1S2624 (Cormand et al., 1997b), with no recombination detected between the gene and markers D1S2777, D1S2721, and D1S2140.

We previously analyzed several known GD mutations in 35 Spanish patients (Cormand et al., 1995). In this paper, we present the molecular analysis of the *GBA* gene in 53 patients. This approach allowed the identification of 95% of the mutant alleles, resulting in the detection of 24 different mutations, 11 described for the first time. Clinical findings were also studied to explore the genotype–phenotype correlations. In addition, linkage disequilibrium between the most common mutations and an intragenic polymorphic site was evaluated and flanking STR markers were used to establish a putative ancestral haplotype, in which the common N370S mutation could have first occurred.

## MATERIALS AND METHODS

### Patients

Fifty-three unrelated Spanish GD patients with different clinical subtypes (42 type I, 8 type II, and 3 type III) were studied. They originate from different locations in continental Spain and the Canary islands. A clinical evaluation was provided by the patient's physician, in most cases including physical examination, routine hematological and laboratory tests, and radiological and neurological studies. In all cases, the GD diagnosis was confirmed by demonstration of low  $\beta$ -glucocerebrosidase activity in leukocytes or cultured fibroblasts. Some of these patients are described elsewhere (Chabás et al., 1995, 1996; Cormand et al., 1995, 1996). Several Argentinian patients (submitted) were included for the analysis of haplotypes associated with mutation N370S. Appropriate informed consent was obtained from all patients.

### Enzymatic Analyses

The  $\beta$ -glucosidase activity was measured with N-stearoyldihydroglucosylceramide (1 mM) or 4-methylumbelliferil- $\beta$ -glucopyranoside (4.5 mM) in the presence of sodium taurocholate (1.5% w/v) and Triton X-100 (0.2% v/v), as previously reported (Cormand et al., 1995).

### DNA Isolation

Genomic DNA was prepared from harvested skin fibroblasts, peripheral blood leukocytes, or spleen, using a standard method (Miller et al., 1988).

### Detection of Nine Previously Described *GBA* Mutations

All patients were screened for eight known mutations in the *GBA* gene (N370S, L444P [1448T→C], D409H [1342G→C], R463C [1504C→T]), 84GG, IVS2+1, RecTL and RecNciI). The genomic samples were amplified by polymerase chain reaction (PCR) and examined by either allele-specific oligonucleotide (ASO) hybridization or restriction analysis (Cormand et al., 1995). The gene/pseudogene recombinant alleles were analyzed by determining the presence of the point mutations L444P (1448T→C), A456P (1483G→C) and V460V (1497G→C) for RecNciI, and D409H (1342G→C), L444P (1448T→C), A456P (1483G→C), and V460V (1497G→C) for RecTL. In addition, the 1263del55 mutation was screened by gene-specific PCR amplification of a 476-bp fragment [sense primer, nt 5569–5588; antisense primer, nt 6025–6044; sequence numbering according to Horowitz et al. (1989)]; and separation of the products on a 2% agarose gel.

### PCR Amplification and SSCP Analysis

PCR and single-strand conformation polymorphism (SSCP) analysis of 14 DNA fragments covering the 11 exons of the *GBA* gene, the flanking intronic junctions, part of the promoter region, including the TATA box, and the first polyadenylation signal, was performed from genomic DNA of the patients. The size of the PCR fragments used ranged from 139 bp to 292 bp, in order to achieve optimal results in the SSCP analysis (Orita et al., 1989). All primer pairs were chosen to amplify the gene, and not the highly homologous *GBA* pseudogene, except in exon 11, for which this was not possible. For gene-specific amplification of this last exon, a nested PCR was performed from a larger, gene-specific PCR fragment. Information on primers is available from the authors.

Although many primers presented several mismatches with the pseudogene sequence, in six of the pairs one of the primers was nonspecific while the other presented only one mismatch in the last nucleotide (3'), in relation to the pseudogene sequence. In these cases, the G/T heteroduplex was avoided, as it is relatively stable (Kwok et al., 1994). Nevertheless, before the systematic mutation scanning, each of the 14 fragments was amplified on genomic DNA from a normal individual, cloned, and sequenced. The sequences of 15 clones from each product showed that, in all cases, only the gene, and not the pseudogene, was amplified.

For all fragments, the PCR reaction was performed under the following conditions: 100 ng of template

DNA, 1 U of Dynazyme DNA polymerase (Finnzymes Oy, Finland), 200  $\mu$ M dNTPs, 20 pmol of each primer, in the recommended buffer in a final volume of 50  $\mu$ l. The PCR program consisted of 35 cycles of denaturation at 94°C for 30 sec and a single annealing/extension step at 55°C for 30 sec.

SSCP analysis was performed as described in Cormand et al. (1997a). Four SSCP conditions were tested for each fragment, combining different polyacrylamide concentrations (8% or 12% acrylamide–bisacrylamide 29:1), different glycerol concentrations in the gel (0% or 5% glycerol), and two running conditions (RT at 200 V, or 4°C at 300 V, always 16 hr).

### Cloning and Sequencing

Samples showing abnormal SSCP patterns were reamplified from genomic DNA and the new PCR reactions were subsequently purified by Wizard™ PCR Preps (Promega, Madison, WI) and cloned into pUC18 vector using the SureClone™ Ligation Kit (Pharmacia Biotech, Gaithersburg, MO). In each case, 6–10 clones were sequenced by fluorescent dideoxy cycle sequencing (ABI 373A Fluorescent DNA Sequencer, Perkin Elmer, Norwalk, CT). Alternatively, in four cases, to analyse each allele independently avoiding the time-consuming subcloning steps, one of the abnormal SSCP bands was cut out of the gel after silver staining and eluted in 50  $\mu$ l of water for 2 hr at 50°C; 20  $\mu$ l was used in a 100- $\mu$ l PCR reaction. The PCR fragments were column-purified and sequenced.

Mutations P391L (–*Ava*II), G389E (+*Mbo*II), R359Q (–*Taq*I), Y313H (–*Kpn*I), G202R (–*Msp*I), G195E (+*Hinf*I), R120W (–*Msp*I), and 1451delAC (–*Hga*I) were confirmed by digestion of the corresponding PCR products with the indicated enzyme. The remaining rare GD alleles were confirmed by direct sequencing of PCR-amplified fragments.

### Protein Secondary Structure

The secondary structure for the normal and mutant proteins were predicted by the method of Chou and Fasman (1978), using the GCG package (Devereux et al., 1984).

### Analysis of the 6144A→G GBA Polymorphism

The 6144A→G polymorphism in intron 9 of the *GBA* gene was analyzed by *Hha*I digestion of a PCR-amplified product. Gene-specific primers (sense: nt 5904–5923, antisense: 6655–6690) were used to amplify a 787-bp fragment (Sidransky et al., 1992). The alleles were designated + (*Hha*I site present) or – (*Hha*I site absent), according to the method of Beutler et al. (1992).

The association of the N370S and L444P alleles with either the + or the – variants of the intragenic 6144A→G polymorphism was studied. In double heterozygous patients, phase was established either by genotyping the parents or by cloning a DNA fragment containing both the mutation and the polymorphic site.

### Haplotype Analysis

Several microsatellite markers (D1S2140, D1S2777, D1S2721, D1S2624, and D1S305) (Fig. 1) located close to the *GBA* gene (Cormand et al., 1997b) were used to genotype 24 N370S chromosomes. These markers belong to the MapPairs set (Research Genetics, Huntsville, AL) and were analyzed according to the manufacturer's recommendations.

### Statistical Analyses

Allelic associations between the diallelic 6144A→G polymorphism and the common N370S and L444P mutations were evaluated using the  $D' = D/D_{\max}$  standardized linkage disequilibrium coef-

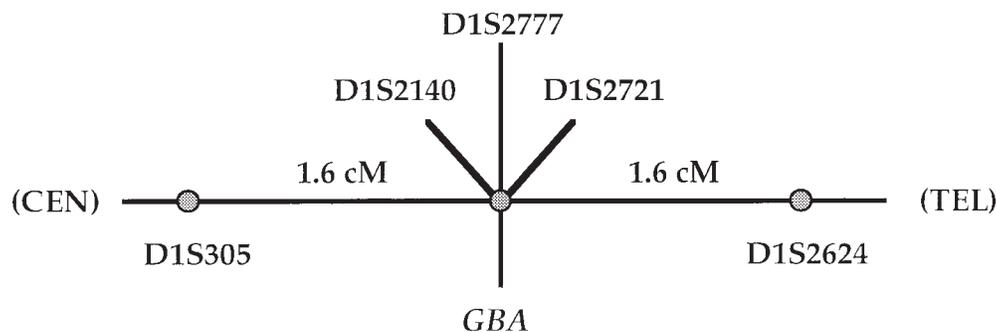


FIGURE 1. Genetic map of microsatellite markers used for linkage disequilibrium and haplotype analysis at 1q21. The genomic region spanned by D1S305 at the centromeric boundary and D1S2624 at the telomeric boundary with the

corresponding genetic distances (Dib et al., 1996) is shown schematically. The glucocerebrosidase locus (*GBA*) was recently mapped by the authors close to markers D1S2140, D1S2777, and D1S2721 (Cormand et al., 1997b).

ficient (Lewontin, 1988). The significance of association was tested by a one-sided  $\chi^2$  test.

For markers with multiple alleles, saturated hierarchical log-linear models were fitted to the data (Fañanás et al., 1992; E. Monrós and J. Bertranpetit, unpublished data). The  $\lambda$  interaction parameter is an estimate of the degree of association between each allele of a marker and the presence or absence of the disease mutation. The significance of this association was measured through the normalized value ( $z$ ) of  $\lambda$ , obtained by dividing lambda by its standard error. An association was considered significant at the 0.05 level if  $|z| > 1.96$ . Because of the large number of comparisons undertaken, the Bonferroni correction according to Rice (1989) has been used. The SPSS statistical package version 6.01 was used to fit log-linear models.

The homozygosity ( $j$ ) at the GBA locus in a given population was determined by

$$j = \sum x_i^2$$

where  $x_i$  is the frequency of the  $i$ th allele (Guldberg et al., 1996). As the identification of GD mutations is not 100% in most populations studied, we have calculated maximum and minimum values of the homozygosity index, either considering all the unidentified alleles as a single one or defining the unknown alleles as having a frequency of  $1/N$ , where  $N$  is the total number of mutant chromosomes studied.

## RESULTS

### Clinical Manifestations

The clinical characteristics of GD among 42 type I patients are summarized in Table 1. Patients are grouped according to genotype. The age of presentation is shown for all patients (if known), while the age at evaluation is only indicated for those patients whose age at first symptom presentation was unknown or for those with early manifestations of the disease. Phenotypes of patients I.1–I.26, documented previously (Cormand et al., 1995), are listed for comparison with the remaining patients and because additional clinical signs are indicated in some cases.

Splenomegaly was noted in all type I patients, except patient I.26, although his affected sibling presented with hepatosplenomegaly. At the age of 40, patient I.26 had been diagnosed of an atypical parkinsonian syndrome because of hypokinesia, rigidity and tremors affecting mainly the right side, and a poor response to conventional anti-parkinson therapy. Two patients with early manifestations of GD in childhood (I.31 and I.38) had short stature, a sign that has been associated with

massive spleen enlargement in children. Hepatomegaly was present in 23 out of the 42 type I patients (54.7%). Liver disease was encountered in patients I.4, I.27, and I.31 and was probably cause of death in patient I.32. Various degrees of bone involvement were found in 13 type I patients (31%). They showed widening of the distal femurs (Erlenmeyer flasks deformity) or hip osteoarthritis as the most frequent signs. In type I GD patients, thrombocytopenia was the most common cytopenia (15 patients, 35.7%); 5 patients had anemia and 9 patients had pancytopenia.

Table 2 shows the clinical findings of 11 patients with the neuronopathic forms of the disease. The time of onset ranged from birth to age 11 months in type II patients. Case II.2 was a fetus that was aborted following prenatal diagnosis. The parents had previously had an affected child, who was diagnosed for the first time at the age of 8 months with persistent emesis and rectal bleeding. Significant clinical signs were hepatosplenomegaly with abdominal pain, generalized eczema, and moderate thrombocytopenia. Microcephaly, deafness, and spasticity were also observed. Massive bleeding from birth was also present in patients II.7 and II.8. Case III.3 is the only individual alive among the type III patients so far analyzed. His neurological condition has not further deteriorated at the present age of 13, while splenomegaly has progressed with severe thrombocytopenia.

### Mutation Analysis

We have analyzed 106 Spanish GD chromosomes by digestion with restriction enzymes, ASO hybridization, SSCP analysis, and sequencing, and we have identified 24 different mutations in the GBA gene that account for 95.3% of all Spanish GD alleles.

All the mutations identified and their genomic localization are shown in Figure 2. The genotypes of the patients included in this study are given in Tables 1 and 2.

Two amino acid substitutions, N370S and L444P, account for 66.1% of the total number of alleles, while the other 22 mutations are represented at frequencies lower than 5%. Mutation D409H (4.7%) is the third most common GD allele in this group of patients (Table 3). Three patients carry the RecNciI mutation, a crossover between the GBA gene and its closely linked pseudogene (Zimran et al., 1990b). This systematic screening revealed 14 novel mutations, including 8 amino acid substitutions (G113E [455G→A], R120W [475C→T], T134P [517A→C], Y313H [1054T→C], G389E [1283G→A], P391L [1289C→T], N392I [1292A→T], Y412H [1351T→C]), 3 stop codons (W(-4)X [108G→A],

TABLE 1. Patients and Genotypes: Type I Patients

Patient subtype/No.	Genotype	Age <sup>a</sup>	Clinical signs
I.2, I.7, I.9, I.12, I.17, I.36	N370S/N370S	27, 68, 4, 56, 58, 82	All S but not H; three asymptomatic; most very mild
I.4, I.10, I.14, I.16, I.19, I.20, I.21, I.23, I.25, I.34, I.37, I.39	N370S/L444P	2 (32), 33, 29, 25, 19, 48, 7 (31), 5, 24, inf (23), 60, 3	All S. 6 patients also H. 7 with pancytopenia, 4 with bone disease; one (I.39) asymptomatic at present age of 5 yr
I.13, I.30, I.31	N370S/RecNcil	28, 42, 13	I.13: SH, thrombocytopenia, cystalgia; I.30: S, thrombocytopenia, leukopenia, chronic diarrhea, skin pigmentation; I.31: SH, thrombocytopenia, leukopenia, stature and pubertal development delayed; portal hypertension; irregularity of right humerus
I.8, I.24	N370S/P391L	31, 27	Both: SH, epistaxis; I.8: pancytopenia, femur deformity. I.24: thrombocytopenia
I.1	N370S/?	7	S only
I.3	N370S/?	3	SH; macrocephaly
I.5	N370S/ 1451delAC	15	S, thrombocytopenia, anemia, bone pain
I.6	N370S/Y313H	10	SH, thrombocytopenia
I.11	N370S/1098insA	3 (9)	SH, leukopenia, splenectomy at 9 yr
I.15	N370S/R359Q	35	SH, epistaxis, easy bruising, femur deformity, conjunctival pingueculae
I.18	N370S/1263del55	5	SH, femur infarcts
I.26	N370S/500insT	47	No SH; atypical parkinson <sup>b</sup>
I.27	N370S/T134P	? (39)	SH, digestive hemorrhagia, ulcer gastroduodenal, hepatopathy, skin pigmentation, bone deformities
I.28	N370S/W(-4)X	4	SH, mild pancytopenia, epistaxis, bone crisis, femur deformity, skin pigmentation
I.29	N370S/G113E	7 m (4)	SH, anemia, bone disease
I.33	N370S/G202R	25	S, pancytopenia, bone deformity, aseptic necrosis
I.35	N370S/IVS5+1	? (24)	Data not available
I.38	N370S/G195E	8 (36)	SH, anemia, thrombocytopenia, stature development delayed, jaundice, skin pigmentation
I.40	N370S/Y412H	? (26)	S, thrombocytopenia, familial factor IX deficiency <sup>c</sup>
I.41	N370S/R257X	6	SH, pancytopenia
I.42	N370S/Q169X	? (87)	SH, epistaxis, pancytopenia
I.22	D409H/E326K	2.5	SH only (present age 14 yr)
I.32	??	? (76)	SH; hepatopathy, cardiomegaly, hip osteoarthritis, vitiligo, pale skin

S, splenomegaly; H, hepatomegaly; inf, infancy.

<sup>a</sup>At presentation (years). Age at evaluation is indicated in parentheses for patients with an early or unknown age of disease presentation.

<sup>b</sup>Affected brother (46 yr) with SH, thrombocytopenia, and pingueculae.

<sup>c</sup>Affected sibling (23 yr) with same symptoms plus bone disease.

Q169X [622C→T], R257X [886C→T]), two 1-bp insertion (500insT, 1098insA), and one donor splice-site mutation (IVS5+1G→T). Data on mutations R120W, Y313H, and 1098insA are reported elsewhere (Chabás et al., 1996; Cormand et al., 1996), and the rare G202R (721G→A), E326K (1093G→A), and R359Q (1193G→A) have already been detected in non-Spanish patients once before (Kawame et al., 1992; Beutler et al., 1994). Very recently, two of the mutations detected in this work, G195E [701 G→A] and 1451delAC, were described for the first time (Grace et al., 1997). It should be noted that the Spanish patient bearing the 1451delAC mutation, described by these authors, is our patient I.5 who received treatment at Mt. Sinai Hospital. Most of these new or rare mutations were found only once

in this panel of patients. Only mutations P391L and G195E were found twice.

Screening of 80 normal chromosomes failed to detect any of the 24 mutations described in this study. After this systematic screening, five GD alleles remained unidentified. Three of them were present in heterozygosis with a known GD allele in patients I.1, I.3, and I.4 (Tables 1, 2), whereas no mutation was identified in patient I.32. Patient I.3 was initially genotyped as N370S/N370S in a previous study (Cormand et al., 1995), but further analysis of the parents showed that only the mother was a carrier of the mutation. Eight polymorphic microsatellite markers in different chromosomes were used to confirm paternity, and Southern blot analysis after *Bam*HI digestion (data not shown) showed a band pattern different from that of wild-type individuals, indicat-

TABLE 2. Patients and Genotypes: Type II and III Patients

Patient subtype/No.	Genotype	Age at presentation/death	Clinical signs
II.1	L444P/L444P	1 m/8.5m	SH, thrombocytopenia, anemia, seizures, head retroflexion, hyperreflexia
II.7	L444P/L444P	2 m/3m	SH, growth retardation, thrombocytopenia, massive bleeding, coagulopathy, bronchoaspiration
II.8	L444P/L444P	birth/2.5 m	SH, low-weight neonate, anemia, thrombocytopenia, rectal bleeding, ascites, hepatocellular damage, cholestatic jaundice
II.2	L444P/G195E	Fetus	
II.3	L444P/R120W	5 m/9 m	SH, generalized hypertonia, neck hyperextension, opisthotonos, strabismus, osteoporosis, pulmonary fibrosis
II.4	G389E/?	6 m/10 m	SH, myoclonic seizures, hypertonia, psychomotor retardation
II.5	L444P/N392I	6 m/12 m	SH, hypertonia, opisthotonos, strabismus, head retroflexion
II.6	D409H/R120W	11 m/20 m	SH, generalized spasticity, psychomotor retardation
III.1	L444P/L444P	3 yr/7.5 yr	SH, respiratory complications, bone pain, and fractures
III.2	D409H/D409H	?/19 yr	SH, epistaxis, strabismus, ophthalmoplegia, pes cavus, cardiovascular disease <sup>a</sup>
III.3	L444P/D409H	3 yr	SH, strabismus, ophthalmoplegia, thrombocytopenia (present age: 13 yr)

S, splenomegaly; H, hepatomegaly.

<sup>a</sup>One affected sibling without H (see Chabás et al., 1995, for further details).

ing that a deletion, or another type of rearrangement, could be the cause of this erroneous assignment, similar to the case described by Beutler and Gelbart (1994). Further analyses on this case are currently being undertaken.

#### Heterogeneity in the GBA Gene in Different Populations

In order to compare the mutation heterogeneity at the GBA locus between the Spanish and other populations, we used allele frequencies to calculate the homozygosity as in Gulberg et al. (1996) (Table 3). We chose those populations for which at least 70% of the mutated alleles were identified. Norrbottnian GD patients are all homozygous for the same mutation (L444P) (Dahl et al., 1990), giving a homozygosity value of 1. Ashkenazi Jewish patients are also quite homogeneous (0.50–0.59) due to the high frequency of mutations N370S and 84GG (Horowitz et al., 1993; Balicki and Beutler, 1995). The Spanish disease population has an intermediate value of 0.24, higher than most non-Jewish GD populations.

#### Linkage Disequilibrium Between the 6144A→G Intragenic Polymorphism and Mutations N370S and L444P

We analyzed the 6144A→G intragenic polymorphism in intron 9 of the GBA gene in all 53 Spanish GD patients. In addition, 44 nonaffected Spanish

individuals were genotyped in order to determine the frequency of the two alleles in the Spanish population. The + allele accounts for 29.5% (26 of 88) of the chromosomes studied and the – allele for 70.5% (62 of 88), in agreement with previous reports (Sorge et al., 1985; Glenn et al., 1994). As the 12 polymorphic sites described within the GBA gene have been reported to be in complete linkage disequilibrium (Beutler et al., 1992b), we analyzed only one of them, the 6144A→G (*Hha*I) site.

Our data revealed that mutation N370S was always associated with the common – variant in these patients (46 of 46). Mutation L444P was found associated with both the + (15 of 24) and the – (9 of 24) alleles. However, there was a clear overrepresentation of L444P alleles in the context of the less frequent + allele. Significant linkage disequilibrium values were obtained in both cases (N370S:  $D = 0.067$ ,  $D' = 1$ ,  $P < 0.001$ ; L444P:  $D = 0.055$ ,  $D' = 0.408$ ,  $P < 0.005$ ).

#### Haplotype Analysis of N370S Chromosomes

In order to study identity by descent by shared haplotype analysis, five microsatellite markers closely linked to the GBA gene on chromosome 1q21 were analyzed in 24 chromosomes bearing the common N370S mutation. We studied 6 unrelated Spanish GD patients with genotype N370S/N370S (I.2, I.7, I.9, I.12, I.17, and I.36) and 3 heterozygous for N370S

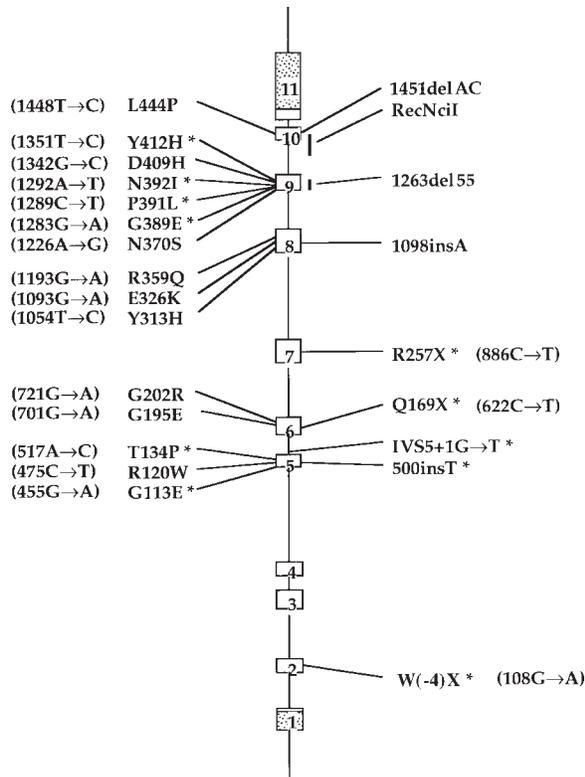


FIGURE 2. Schematic representation of the mutations identified in the  $\beta$ -glucocerebrosidase gene in Spanish Gaucher disease patients. Missense mutations are depicted on the left; gene/pseudogene recombinations, insertions, deletions, aberrant splice sites, and nonsense mutations are shown on the right. Novel mutations are indicated by an asterisk (\*). In brackets, the cDNA position of the nucleotide change; shaded boxes, untranslated regions, and numbers correspond to exons.

(I.6, I.11, and I.18). Additionally, we analyzed 9 chromosomes, bearing the N370S allele, from Argentinian patients. Three of these chromosomes were found in two patients of Ashkenazi Jewish origin.

Full haplotypes for the five loci were established. One major haplotype, 1-4-3, for markers D1S2777, D1S2140, and D1S2721, respectively, was present in 16 of 24 (66.7%) of the N370S chromosomes analyzed (Table 4). Eight chromosomes, belonging to individuals I.6, I.11, I.18, I.36, and to three of the Argentinian patients, bore other allelic combinations: 1-4-7 (I.6), 1-4-6 (I.11), 1-4-1 (I.18), 6/1-1/4-2/1 (I.36, phase unknown) and 1-3-8, 1-6-1, 1-4-1 (Argentinian patients). Interestingly, the Spanish consensus haplotype is identical to the haplotype present in the 3 chromosomes of known Ashkenazi Jewish origin.

In order to measure the degree of allelic association between each marker and the N370S mutation, we applied hierarchical log-linear models. Table 4 summarizes the results of this analysis. Significant values of association ( $|z| > 1.96$ ) were obtained for

alleles 1, 4, 3, and 5 of markers D1S2777, D1S2140, D1S2721, and D1S2624, respectively.

## DISCUSSION

### Mutation Analysis

The prevalence of GD is at its highest in the Ashkenazi Jewish population, thus most studies have focused on this ethnic group. Most of the studies of mutations causing GD carried out on non-Jewish patients were performed on individuals of diverse ethnic (Beutler and Gelbart, 1993) or nonspecified origin (Horowitz et al., 1993). By contrast, the studies on specific non-Jewish populations were limited to the analysis of a relatively small group of previously described mutations (Walley et al., 1993; Cormand et al., 1995; Michelakakis et al., 1995; Amaral et al., 1996; Tylki-Szymanska et al., 1996; le Coutre et al., 1997). Here we report an exhaustive search for the mutations present among Spanish GD patients, which allowed the identification of 95% of the mutant alleles in the 53 patients analysed. Mutations N370S, L444P, D409H, and RecNciI account for around 73% of the total number of GD chromosomes, while the remaining 27% is accounted for by at least 20 different mutations, including 14 novel, three of them already reported by us (Chabás et al., 1996; Cormand et al., 1996). After extensive SSCP screening covering the 11 exons of the *GBA* gene, all the splice sites and part of the 5' and 3' untranslated regions, only 5 mutant alleles remained unidentified.

Five of the 11 novel variants described in this paper are undoubtedly disease-causing mutations, as they create a frameshift (500insT), premature stop codons (W(-4)X, Q169X, R257X), or give rise to abnormal splicing (IVS5+1G→T). Mutation IVS5+1G→T was the only noncoding alteration found. Although the effect of this substitution in the first nucleotide of intron 5 was not studied in the patient RNA because no sample was available, this position is 100% conserved in the vertebrate donor splice-site consensus (Padgett et al., 1986). The other 6 new mutations are amino acid substitutions, and several findings strongly suggest that they are also disease-causing mutations: (1) after exhaustive examination of the *GBA* gene, no other mutation was found in any of the patients; (2) the changes were not present in 80 normal chromosomes analyzed; (3) all the amino acid residues involved are conserved in murine and human *GBA*, indicating functional/structural relevance; and (4) prediction of protein secondary structure indicates that all mutations except for Y412H may produce conformational changes in the mature enzyme. Moreover, all these

TABLE 3. Prevalence (%) of the Four Most Common GBA Mutations in Different Populations and Homozygosity ( $\Sigma x_i^2$ ) at the GBA Locus as a Measure of Genetic Heterogeneity

Population	$\Sigma x_i^2$ <sup>a</sup>	N370S	84GG	IVS2+1	L444P	D409H	RecNcil <sup>b</sup>	R463C	G377S	No. of chromosomes	% mutation detection
Norrbottnian <sup>1</sup>	1.00 (1.00)	—	—	—	100	—	—	—	—	22	100
Ashkenazi Jewish <sup>2</sup>	0.59 (0.59)	76.45	11.96	2.17	3.26	—	—	—	—	276	99.3
Ashkenazi Jewish <sup>3</sup>	0.50 (0.51)	69.77	10.17	2.26	4.24	—	—	—	—	354	90.4
Portuguese <sup>4</sup>	0.31 (0.33)	53.70	—	—	12.96	—	—	3.70	7.41	54	87.0
Spanish <sup>5</sup>	0.24 (0.25)	43.40	—	—	22.64	4.72	2.83	—	—	106	95.3
Non-Jewish <sup>3</sup>	0.19 (0.20)	22.86	—	—	31.43	5.00	7.80	—	—	240	89.2
Australasian <sup>6,7</sup>	0.17 (0.25)	25.00	3.57	—	30.36	—	7.14	—	—	56	69.6
Non-Jewish <sup>2</sup>	0.16 (0.21)	30.00	—	—	30.00	—	4.17	4.17	—	140	77.1

<sup>a</sup>Less common and uncharacterized alleles were counted as different alleles or as one single allele (the latter shown in brackets).

<sup>b</sup>Including cases described as XQVR.

References: <sup>1</sup>Dahl et al. (1990); <sup>2</sup>Balicki and Beutler (1995); <sup>3</sup>Horowitz et al. (1993); <sup>4</sup>Amaral et al. (1996); <sup>5</sup>present study; <sup>6</sup>Lewis et al. (1994); <sup>7</sup>Nelson et al. (1995).

TABLE 4. Allelic Association Between the N370S Mutation and Markers on Chromosome 1q21

Marker	No. of alleles/ heterozygosity <sup>a</sup>	Associated allele		% of N370S chromosomes	% of control chromosomes <sup>c</sup>	$\lambda$	z	P*
		Designation <sup>b</sup>	Sequence					
D1S305	9/0.86	7	(AC) <sub>17</sub>	12.5% (3/24)	3.6% (2/56)	0.70	1.72	<0.09
D1S2777	9/0.57	1	(AC) <sub>16</sub>	95.8% (23/24)	57.1% (32/56)	0.57	2.13	<0.04
D1S2140	7/0.71	4	(AT) <sub>6...</sub> (TATC) <sub>11</sub>	87.5% (21/24)	30.4% (17/56)	0.83	3.16	<0.002
D1S2721	8/0.65	3	(AC) <sub>18</sub>	66.7% (16/24)	23.1% (12/52)	0.70	2.69	<0.008
D1S2624	5/0.71	5	(AC) <sub>20</sub>	16.7% (4/24)	1.8% (1/56)	0.92	2.24	<0.03

<sup>a</sup>Dib et al. (1996).

<sup>b</sup>According to CEPH-Généthon nomenclature.

<sup>c</sup>CEPH-Généthon data.

\*All values are significant at  $P < 0.05$ , except for allele 7 of marker D1S305.

substitutions involve a change in the polarity or charge of the residue (G113E, T134P, G389E, N392I, Y412H) or are structurally disruptive (P391L). However, further characterization of the products expressed from each mutant allele is needed to confirm these assumptions.

Our results show that the 15 mutations leading to amino acid substitutions are mostly confined to a few exons, while nonsense, frameshift and splice-site mutations are widely distributed throughout the gene (Fig. 2). The clustering of naturally occurring missense mutations in exons 8 and 9 (9 of 15) and in exons 5 and 6 (5 of 15) of the *GBA* gene is consistent with previous data (Beutler and Gelbart, 1996), and confirms the important structural/functional roles of these regions, as deduced from biochemical analyses (Grace et al., 1990, 1994; Ohashi et al., 1991).

In order to enhance the sensitivity of the SSCP method, we assayed a battery of different conditions such as polyacrylamide and glycerol concentration in the gel and running temperature, and we chose the best four combinations for each amplified sequence. The importance of varying the experimental parameters is demonstrated by the finding that some of the mutations described here were detected in only one or two of the four conditions assayed. The intrinsic limitations of the SSCP technique or the presence of mutations outside the regions analyzed may explain why four mutated alleles were not detected (the fifth missing allele is the uncharacterized deletion of patient I.3). Complete sequencing of the *GBA* gene or the cDNA from patients bearing these unidentified alleles may reveal the underlying mutations. Interestingly, patient I.32, with clear symptoms of GD, is the only one with two unidentified mutations, and this leaves open the unlikely possibility of a different genetic cause for the disease. The low  $\beta$ -glucocerebrosidase activity rules out the prosaposin gene as responsible for GD.

The complete mutation analysis performed on the

Spanish GD patients provides interesting data for molecular diagnosis. The detection of the most prevalent mutations in this population, N370S, L444P, D409H, and RecNciI, allows full identification of nearly half of the genotypes (47.2%). Moreover, another 35.8% is partially covered by this analysis, as they are heterozygous for N370S, which provides essential information for the prognosis of the disease.

#### Genotype/Phenotype Correlations

The large heterogeneity observed in clinical manifestations of GD results, in part, from the large number of mutations in the *GBA* gene. However, many unrelated or related patients with the same genotype have been described with significantly distinct clinical features, indicating that other factors may play an important role.

The presence of the N370S mutation, either in homozygosity or in heterozygosity, always precludes development of neurological manifestations in our patients, in agreement with previous data (Beutler and Grabowski, 1995). However, the phenotypic expression is highly variable. Those with genotype N370S/N370S are the least severe, including three asymptomatic individuals (see Table 1). The age of onset varies within this group from 4 to 82 years. Patients with genotype N370S/L444P present more severe symptoms and the average age of onset ( $23.2 \pm 18.0$  years) is significantly lower than that of patients who are homozygous for N370S ( $49.2 \pm 26.1$ ). Phenotypic variation is also observed among the three patients with genotype N370S/recNciI and between the two patients with N370S/P391L.

It is well established that homozygosity for the L444P mutation causes severe disease with neurological involvement, but the clinical expression ranges from the juvenile form or type III to the infantile neuronopathic form or type II (Beutler, 1995). Interestingly, some Japanese L444P/L444P patients have been found among the non-neuronopathic form of

GD (Ida et al., 1995). Four unrelated Spanish GD patients have been genotyped as L444P/L444P (Table 2). Three of these patients represent the type II of the disease, and the other was classified as type III.

Recently the association between the D409H/D409H genotype and a unique clinical presentation consisting of cardiovascular abnormalities and oculomotor apraxia (Abrahamov et al., 1995; Uyama et al., 1997) has been reported. This association is illustrated in our group of patients by individual III.2 and her two sisters, described in detail elsewhere (Chabás et al., 1995).

Only two of the 24 mutations described in this study could be classified as mild: the most prevalent N370S and E326K. Mutations 1451delAC, 1263-del55, 500insT, 1098insA, W(-4)X, Q169X, R257X, and IVS5G→T belong to the "lethal" group, and mutations R120W, G195E, G389E, N392I, D409H, and L444P to the "severe" group, according to the classification by Beutler et al. (1994). Eight mutations could not be assigned to any group because they were found together with the common mild N370S mutation, which is always associated with a type I phenotype.

#### Linkage Disequilibrium, Haplotype, and Population Analyses

From our data it is reasonable to conclude that the L444P mutation arose more than once in this group of patients, since it appears in the context of both the + and - variants of the intragenic polymorphism. It is significant that this mutation is present in the normal pseudogene sequence, suggesting that gene conversion events could be responsible for these recurrent changes. However, significant linkage disequilibrium was observed between this mutation and the less common + allele ( $D = 0.055$ ,  $D' = 0.408$ ,  $P < 0.005$ ), suggesting expansion of an ancestral allele.

The N370S allele is invariably associated with the - allele in our group of patients, in agreement with previous data from the Ashkenazi Jews (Beutler et al., 1992b) and from other populations (Amaral et al., 1996). These results are consistent with a common origin for the panethnic N370S GBA allele. However, because the - allele is about twice as common in the general population, as the + (Zimran et al., 1990a), recurrent mutations cannot be formally ruled out from these data.

We have recently mapped the GBA gene genetically (Cormand et al., 1997b). The analysis of three microsatellite markers, known to be very close to the gene, in 24 N370S alleles from Spanish and Argentinian patients, showed that one major haplotype was present in 66.7% of the chromosomes. The same haplotype was found in the three Ashkenazi Jewish

chromosomes analyzed bearing N370S. It would be interesting to study more chromosomes from the Jewish population to evaluate the possibility of a common origin for the Spanish and the Ashkenazi Jewish N370S mutation.

#### ACKNOWLEDGMENTS

We thank all the physicians who referred the patients, as well as the patients and their families, for their collaboration. The excellent technical assistance of J. Jarque and H. Sellés is also acknowledged. The authors are also grateful to M. Bayés, S. Balcells, R. González-Duarte, F. Palau, S. Bort, E. Monrós, and J. Bertranpetit for their helpful contribution on different aspects of this work, and to R. Rycroft for revising the English. Bru Cormand was a recipient of a fellowship from the CIRIT (Generalitat de Catalunya).

#### REFERENCES

- Abrahamov A, Elstein D, Grosstsur V, Farber B, Glaser Y, Hadashalpern I, Ronen S, Tafakjdi M, Horowitz M, Zimran A (1995) Gaucher's disease variant characterised by progressive calcification of heart valves and unique genotype. *Lancet* 346:1000-1003.
- Amaral O, Pinto E, Fortuna M, Lacerda L, Miranda M (1996) Type 1 Gaucher disease: Identification of N396T and prevalence of glucocerebrosidase mutations in the Portuguese. *Hum Mutat* 8:280-281.
- Balicki D, Beutler E (1995) Gaucher disease. *Medicine (Baltimore)* 74:305-323.
- Beutler E (1995) Gaucher disease. *Adv Genet* 32:17-49.
- Beutler E, Gelbart T (1993) Gaucher disease mutations in non-Jewish patients. *Br J Haematol* 85:401-405.
- Beutler E, Gelbart T (1994) Erroneous assignment of Gaucher disease genotype as a consequence of a complete gene deletion. *Hum Mutat* 4:212-216.
- Beutler E, Gelbart T (1996) Glucocerebrosidase (Gaucher disease). *Hum Mutat* 8:207-213.
- Beutler E, Gelbart T (1997) Hematologically important mutations: Gaucher disease. *Blood Cells Mol Dis* 23:2-7.
- Beutler E, Grabowski GA (1995) Gaucher disease. In Scriver CR, Beaudet AL, Sly WS, Valle D (eds): *The Metabolic and Molecular Bases of Inherited Disease*. 7th Ed. New York: McGraw-Hill, pp 2641-2669.
- Beutler E, Demina A, Gelbart T (1994) Glucocerebrosidase mutations in Gaucher disease. *Mol Med* 1:82-92.
- Beutler E, West C, Gelbart T (1992b) Polymorphisms in the human glucocerebrosidase gene. *Genomics* 12:795-800.
- Chabás A, Cormand B, Balcells S, González-Duarte R, Casanova C, Colomer J, Vilageliu L, Grinberg D (1996) Neuronopathic and non-neuronopathic presentation of Gaucher disease in patients with the third most common mutation (D409H) in Spain. *J Inher Metab Dis* 19:798-800.
- Chabás A, Cormand B, Grinberg D, Burguera JM, Balcells S, Merino JL, Mate I, Sobrino JA, González-Duarte R, Vilagelin L (1995) Unusual expression of Gaucher's disease: Cardiovascular calcifications in three sibs homozygous for the D409H mutation. *J Med Genet* 32:740-742.
- Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequences. *Adv Enzymol* 47:45-158.

- Cormand B, Vilageliu L, Burguera JM, Balcells S, González-Duarte R, Grinberg D, Chabás A (1995) Gaucher disease in Spanish patients: Analysis of eight mutations. *Hum Mutat* 5:303–309.
- Cormand B, Vilageliu L, Balcells S, González-Duarte R, Chabás A, Grinberg D (1996) Two novel (1098insA and Y313H) and one rare (R359Q) mutations detected in exon 8 of the beta-glucocerebrosidase gene in Gaucher's disease patients. *Hum Mutat* 7:272–274.
- Cormand B, Grinberg D, Gort L, Fiumara A, Barone R, Vilageliu L, Chabás A (1997a) Two new mild homozygous mutations in Gaucher disease patients: Clinical signs and biochemical analyses. *Am J Med Genet* 70:437–443.
- Cormand B, Montfort M, Chabás A, Vilageliu L, Grinberg D (1997b) Genetic fine localization of the  $\beta$ -glucocerebrosidase (GBA) and prosaposin (PSAP) genes: Implication for Gaucher disease. *Hum Genet* 100:75–79.
- Dahl N, Lagerstrom M, Erikson A, Pettersson U (1990) Gaucher disease Type III (Norrbottnian type) is caused by a single mutation in exon 10 of the glucocerebrosidase gene. *Am J Hum Genet* 47:275–278.
- Devereux J, Haerberli P, Smithies O (1984) A comprehensive set of sequences analysis programs for the VAX. *Nucleic Acids Res* 12:387–395.
- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseu P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–154.
- Fañanás L, Moral P, Panadero MA, Bertranpetit J (1992) Complementary genetic markers in schizophrenia: C3, BF and C6 polymorphisms. *Hum Hered* 42:162–167.
- Glenn D, Gelbart T, Beutler E (1994) Tight linkage of pyruvate kinase (PKLR) and glucocerebrosidase (GBA) genes. *Hum Genet* 93:635–638.
- Grabowski GA (1993) Gaucher disease—Enzymology, genetics, and treatment. *Adv Hum Genet* 21:377–441.
- Grace ME, Graves PN, Smith FI, Grabowski GA (1990) Analyses of catalytic activity and inhibitor binding of human acid  $\beta$ -glucosidase by site-directed mutagenesis. *J Biol Chem* 265:6827–6835.
- Grace ME, Newman KM, Scheinker V, Berg-Fussman A, Grabowski GA (1994) Analysis of human acid  $\beta$ -glucosidase by site-directed mutagenesis and heterologous expression. *J Biol Chem* 269:2283–2291.
- Grace ME, Desnick RJ, Pastores GM (1997) Identification and expression of acid  $\beta$ -glucosidase mutations causing severe type 1 and neurologic type 2 Gaucher disease in non-Jewish patients. *J Clin Invest* 99:2530–2537.
- Guldberg P, Levy HL, Hanley WB, Koch R, Matalon R, Rouse BM, Trefz F, de la Cruz F, Henriksen KF, Guttler F (1996) Phenylalanine hydroxylase gene mutations in the United States: Report from the Maternal PKU Collaborative Study. *Am J Hum Genet* 59:84–94.
- Horowitz M, Tzuri G, Eyal N, Berebi A, Kolodny EH, Brady RO, Barton NW, Abrahamou A, Zimran A (1993) Prevalence of nine mutations among Jewish and non-Jewish Gaucher disease patients. *Am J Hum Genet* 53:921–930.
- Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E (1989) The human glucocerebrosidase gene and pseudogene: Structure and evolution. *Genomics* 4:87–96.
- Ida H, Iwasawa K, Kawame H, Rennert OM, Maekawa K, Eto Y (1995) Characteristics of gene mutations among 32 unrelated Japanese Gaucher disease patients: Absence of the common Jewish 84GG and 1226G mutations. *Hum Genet* 95:717–720.
- Kawame H, Hasegawa Y, Eto Y, Maekawa K (1992) Rapid identification of mutations in the glucocerebrosidase gene of Gaucher disease patients by analysis of single-strand conformation polymorphisms. *Hum Genet* 90:294–296.
- Kwok S, Chang S-Y, Sninsky JJ, Wang A (1994) A guide to the design and use of mismatched and degenerated primers. *PCR Methods Appl* 3:S39–S47.
- le Coutre P, Demina A, Beutler E, Beck M, Petrides PE (1997) Molecular analysis of Gaucher disease: distribution of eight mutations and the complete gene deletion in 27 patients from Germany. *Hum Genet* 99:816–821.
- Lewis BD, Nelson PV, Robertson EF, Morris CP (1994) Mutation analysis of 28 Gaucher disease patients—The Australasian experience. *Am J Med Genet* 49:218–223.
- Lewontin RC (1988) On measures of gametic disequilibrium. *Genetics* 120:849–852.
- Michelakakis H, Dimitriou E, Vanweely S, Boot RG, Mavridou I, Verhoek M, Aerts J (1995) Characterization of glucocerebrosidase in Greek Gaucher disease patients: Mutation analysis and biochemical studies. *J Inher Metab Dis* 18:609–615.
- Miller SA, Dyke DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215.
- Nelson PV, Carey WF, Morris CP, Lewis BD (1995) Mutation analysis of Australasian Gaucher disease patients. *Am J Med Genet* 58:382.
- Ohashi T, Hong CM, Weiler S, Tomich JM, Aerts JMFG, Tager JM, Barranger JA (1991) Characterization of human glucocerebrosidase from different mutant alleles. *J Biol Chem* 266:3661–3667.
- Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5:874–879.
- Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA (1986) Splicing of messenger RNA precursors. *Annu Rev Biochem* 55:1119–1150.
- Rice WR (1989) Analyzing tables of statistical test. *Evolution* 43:223–225.
- Sibille A, Eng CM, Kim SJ, Pastores G, Grabowski GA (1993) Phenotype/genotype correlations in Gaucher disease type-I—Clinical and therapeutic implications. *Am J Hum Genet* 52:1094–1101.
- Sidransky E, Tsuji S, Martin BM, Stubblefield B, Ginns EI (1992) DNA mutation analysis of Gaucher patients. *Am J Med Genet* 42:331–336.
- Sorge J, Gelbart T, West C, Westwood B, Beutler E (1985) Heterogeneity in Type I Gaucher disease demonstrated by restriction mapping of the gene. *Proc Natl Acad Sci USA* 82:5442–5445.
- Tylki-Szymanska A, Millat G, Maire I, Czartoryska B (1996) Types I and III Gaucher disease in Poland: Incidence of the most common mutations and phenotypic manifestations. *Eur J Hum Genet* 4:334–337.
- Uyama E, Uchino M, Ida H, Eto Y, Owada M (1997) D409H/D409H genotype in Gaucher-like disease. *J Med Genet* 34:175.
- Walley AJ, Barth ML, Ellis L, Fensom AH, Harris A (1993) Gaucher's disease in the United Kingdom—Screening non-jewish patients for the two common mutations. *J Med Genet* 30:280–283.
- Zimran A, Gelbart T, Beutler E (1990a) Linkage of the *sPvuII* polymorphism with the common Jewish mutation for Gaucher disease. *Am J Hum Genet* 46:902–905.
- Zimran A, Sorge J, Gross E, Kubitz M, West C, Beutler E (1990b) A glucocerebrosidase fusion gene in Gaucher disease. Implications for the molecular anatomy, pathogenesis, and diagnosis of this disorder. *J Clin Invest* 85:219–222.