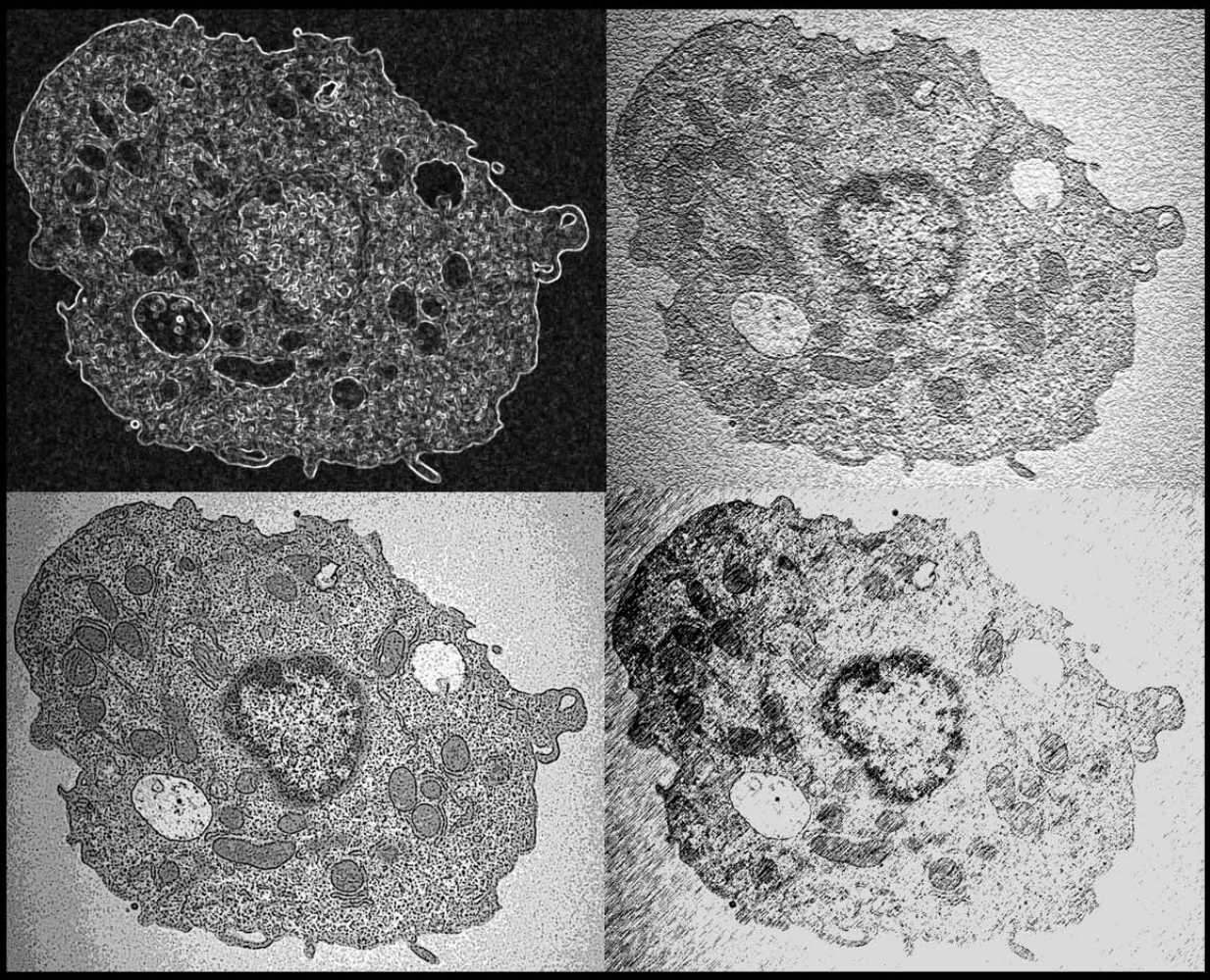
Caracterització de mutacions causants de la malaltia de Gaucher. Aproximació a una teràpia gènica.



Anna Diaz Font

2006

Resultats

CAPíTOL 1:

ANÀLISI I CARACTERITZACIÓ DE MUTACIONS PRODUÏDES PER RECOMBINACIONS ENTRE EL GEN GBA *I EL PSEUDOGÈN CAUSANTS DE LA MALALTIA DE GAUCHER*

1.1. Caracterització de nous al·lels causants de la malaltia de Gaucher produits per recombinacions entre el gen *GBA* i el pseudogen

Un dels objectius d'aquesta tesi era la búsqueda de nous al·lels recombinants que no s'haguessin descrit fins al moment. Disposàvem d'un pacient candidat a ser portador d'una mutació d'aquest tipus. Aquest pacient havia estat diagnosticat com a malalt de Gaucher tipus I i inicialment genotipat com a homozigot per a la mutació N370S. Però aquest pacient presentava una edat d'aparició de la malaltia molt primerenca per ser homozigot per la mutació N370S i en analitzar els pares, que s'esperava que fossin portadors cadascun d'ells de la mutació N370S, es va comprovar que la mare no era portadora d'aquesta mutació.

Una de les explicacions del perquè la mare no portava la mutació i, en canvi, el fill es veiés homozigot per la mutació N370S, era que el fill fos en realitat heterozigot per la mutació N370S i en l'altre al·lel portés una deleció. Aquesta deleció faria que vegèssim la mutació N370S en homozigosi perquè realment només estariem amplificant un al·lel. La deleció hauria d'afectar com a minim la zona d'anellament d'un dels dos encebadors, però podria afectar a tot el gen. Una de les causes que explicaria aquesta deleció podria ser una recombinació entre el gen i el pseudogèn amb la formació d'un gen de fusió entre tots dos.

En analitzar el DNA del pacient vam poder determinar que tenia una deleció que abarcava la majoria del gen com a conseqüència d'una recombinació amb el pseudogèn, a nivell de l'intró 2. Es va poder acotar el punt de la recombinació en una regió de només 18 nucleòtids.

Referència del treball publicat:

A new gene-pseudogene fusion allele due to a recombination in intron 2 of the glucocerebrosidase gene causes Gaucher disease*. Blood Cells, Molecules and Diseases (2000) 26(5): 409-416.*

Cormand et al. Blood Cells, Molecules, and Diseases (2000) 26(5) October: 409–416
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

A New Gene–Pseudogene Fusion Allele Due to a Recombination in Intron 2 of the Glucocerebrosidase Gene Causes Gaucher Disease

Submitted 07/17/00
(Communicated by Ernest Beutler, M.D., 07/17/00)

Bru Cormand,[1,2] Anna Díaz,[1,2] Daniel Grinberg,[1] Amparo Chabás,[3] and Lluïsa Vilageliu[1]

ABSTRACT: Gaucher disease is the most prevalent sphingolipid storage disorder in humans caused by a recessively inherited deficiency of the enzyme glucocerebrosidase. More than 100 mutations have been described in the glucocerebrosidase gene causing Gaucher disease. Some of them are complex alleles with several mutations due to recombination events between the gene and its highly homologous pseudogene. The generation of these recombinant alleles involves, in most cases, a crossover in the 3′ end of the gene, beyond exon 8. However, in a few cases recombination took place in a more upstream location. Here we describe the analysis of a patient with type I Gaucher disease who bears a new complex allele. This allele was originated by a crossover between the gene and the pseudogene at intron 2, the most upstream recombination site described so far, which gave rise to a fusion gene. The patient was first diagnosed as homozygous for the c.1226 A → G (N370S) mutation but the early onset of the disease prompted us to perform parental DNA analysis which showed that the mother was not a N370S carrier, suggesting deletion of at least part of the gene. Molecular analysis of the complex allele was carried out by Southern blot, PCR, and sequencing. We were able to close down the region of the recombination event to an interval of 18 nucleotides, corresponding to the last 15 nucleotides of intron 2 and the first 3 nucleotides of exon 3 of the gene. These 18 nucleotides are identical between the gene and pseudogene making any further refinement impossible. An exhaustive list of published glucocerebrosidase complex alleles, describing their recombination points, is included for comparison.

INTRODUCTION

Gaucher disease (GD) is an inherited deficiency of the lysosomal enzyme glucocerebrosidase, encoded by the *GBA* gene on chromosome 1p21. Many disease-causing alleles have been described in the *GBA* gene (1) most of them being point mutations. A highly homologous pseudogene, *GBAP,* originated by an ancestral gene duplication, is located 16 kb downstream. The presence of these homologous regions in a small genomic interval enhances the probability of rearrangements resulting in the production of the so-called "complex alleles" or "Rec" mutations. These alleles are generated either by crossovers or gene conversion events between the *GBA* gene and its pseudogene. All of them were found in heterozygosis, suggesting lethality in homozygosis. The few homozygous cases described to date correspond to abortions or neonatal deaths (2–4).

The first gene-pseudogene fusion allele was described in 1990 (5). It bears several sequence changes, normally present in the pseudogene: two missense mutations, L444P and A456P, and the silent polymorphism V460V. Since then, alleles with these changes, usually known as Rec*Nci*I, have been reported in different populations, being the most frequent complex allele in GD. Another complex allele, Rec*TL* (6), which bears the same changes as Rec*Nci*I plus D409H, has also been described several times. Now it is clear that each of these Rec alleles could have arisen from different mutational mechanisms such as gene conversion or unequal crossing over. In most cases

Correspondence and reprint requests to: Dra. Lluïsa Vilageliu. Fax: 34-93-411-0969. E-mail: lluisa@porthos.bio.ub.es.
[1] Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, E-08028 Barcelona, Spain.
[2] These authors contributed equally to this work.
[3] Institut de Bioquímica Clínica, Corporació Sanitària, Mejia Lequerica s/n, Edificio Helios III, 08028 Barcelona, Spain.

1079-9796/00 $35.00

Blood Cells, Molecules, and Diseases (2000) **26**(5) October: 409–416 Cormand et al.
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

this issue was not addressed and the name Rec*Nci*I or Rec*TL* is used independently of the mutational mechanism. Besides these two relatively frequent complex mutations, some other recombinant alleles have been reported in individual cases. The recombination site for most of them lies in the 3′ end of the gene, beyond exon 8 (1, 7, 8). Only few crossovers have been described in the 5′ end of the *GBA* gene such as those described by Reissner et al. (9) in intron 3 or Filocamo et al. (10) in intron 6.

Here we report a new recombinant allele, present in heterozygosis in a type I GD patient. An exhaustive molecular characterization of this allele allowed the refinement of the crossover site to an 18-bp interval in the boundary between intron 2 and exon 3 of the *GBA* gene. This is the most upstream recombination event described so far in a GD mutant allele.

MATERIALS AND METHODS

Patient

The patient, now a 14-year-old boy, was clinically diagnosed of Gaucher disease at the age of 2 years 10 months. Diagnosis was confirmed by low glucocerebrosidase activity in cultured fibroblasts. He presented severe splenomegaly from the first months of life but development has been normal despite important splenomegaly (17.5 cm below c.b.) and hepatomegaly. He has no hemorrhagic diathesis. Neurological involvement is absent. Radiologically, the lower end of both femora show the characteristic Erlenmeyer flask appearance. The patient has no bone pain or pathological fractures. Recent analyses show diminished levels of cholesterol and Fe, and normal (in the low range) leukocyte and platelet counts. A bone marrow aspirate showed storage of lipid cells. Clinical data were provided by Dr. T. Toll. (Hospital Sant Joan de Déu, Barcelona, Spain). This patient was partially described as patient I.3 in Cormand et al. (11).

Enzymatic Analyses

The β-glucosidase activity was measured with *N*-stearoyldihydroglucosylceramide (1 mM) or 4-methylumbelliferil-β-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 (12).

DNA Isolation

Genomic DNA was prepared from fibroblasts using the salting out procedure (13).

Sequence Numbering

Nucleotides are numbered throughout the paper according to the recommendations by Beutler and Gelbart (1). The A of the first ATG of the *GBA* gene is considered as nucleotide +1. This implies the subtraction of 583 nucleotides from the genomic sequence by Horowitz et al. (14). This sequence was corrected afterwards (Ref. 15; GenBank Accession No. J03059). However, we have not taken into account this corrected version for nucleotide numbering, in order to allow comparisons with previous published results. This produces small inconsistencies in the length of some of the PCR or Southern fragments.

Detection of the c.1226 A → G (N370S) Mutation

The N370S mutation in exon 9 of the glucocerebrosidase gene was analyzed in the patient through mismatched PCR amplification and *Xho*I digestion as previously described (16). The analysis was also performed on DNA material from the parents and an unaffected brother.

Southern Blot, Probe Preparation, and Hybridization

Genomic DNA was single digested with *Ssp* I, *Bam* HI, and *Xba*I enzymes, electrophoresed on a 0.7% agarose gel, and blotted onto a nylon membrane (Amersham) using standard protocols.

For probe preparation, total RNA was prepared from human cultured fibroblasts by the Ultraspec RNA Isolation System (Biotech). Reverse transcription was performed using the Time Saver cDNA Synthesis Kit (Pharmacia Biotech) with an antisense primer at the 3′-UTR end of the gluco-

Cormand et al. Blood Cells, Molecules, and Diseases (2000) 26(5) October: 409–416
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

cerebrosidase mRNA (5′-CTCTTTAGTCACAGACAGCG-3′, genomic position 6173–6192). The full coding region of the cDNA was PCR-amplified in a 1836-bp fragment using the reverse transcription primer described above and a sense primer at the 5′ end of the cDNA (5′-GCCGGAATTACTTGCAGGGC-3′, genomic position −140 to −121). The PCR was performed with the Expand Long Template PCR System (Boehringer Mannheim). The glucocerebrosidase cDNA probe was labeled with [α-^{32}P]dCTP by random priming.

Southern blot hybridization was carried out at high stringency following a standard protocol. The membrane was exposed to a Hyperfilm-MP (Amersham) for 3 days.

Characterization of the GBA Gene–Pseudogene Fusion by PCR Amplification and Sequencing

A forward gene-specific primer (5′-CCGTGTTCAGTCTCTCCTAG-3′, genomic position 954-973) in intron 2 of the *GBA* gene and a reverse nondiscriminatory primer (5′-CCTCAGGGCCTGAAAAAGCT-3′, genomic position 2601–2620) in intron 5 were used to amplify a 1675-bp gene product in normal chromosomes or a 1024-bp gene/pseudogene fusion product in mutated chromosomes. The PCR was performed under the following conditions: 100 ng of template DNA, 1 U of Expand High Fidelity polymerase (Boehringer Mannheim), 200 μM dNTPs, 10 pmol of each primer, in the recommended buffer in a final volume of 25 μl. The PCR program consisted of 35 cycles of denaturation at 94°C for 40 s, annealing at 62°C for 30 s, and extension at 68°C for 2 min.

The PCR product was purified by GFX PCR DNA and Gel Band purification kit (Amersham Pharmacia Biotech) and sequenced by fluorescence dideoxy cycle sequencing (ABI 373A Fluorescent DNA sequencer, Perkin–Elmer).

RESULTS

Biochemical Analysis

Acid β-glucosidase (substrate 4MU-β-glucoside) and β-glucocerebrosidase (*N*-stearoyl-dihydroglucocerebroside) were 9 and 15% of normal values. Chitotriosidase activity was 16313 nmol/h × ml (215-fold increase over control).

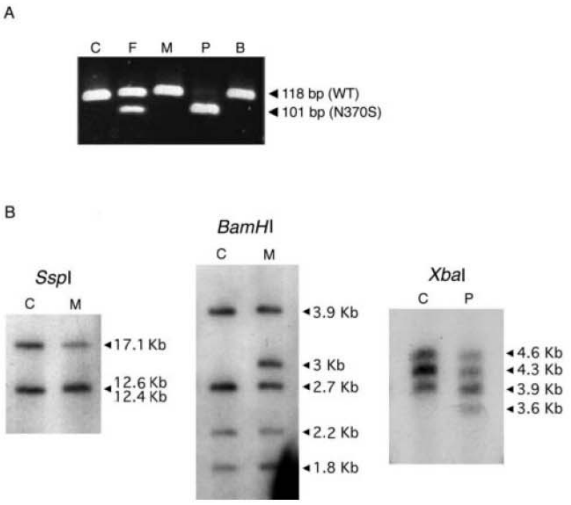


FIG. 1. Characterization of the two mutant alleles of the patient. (A) Detection of mutation N370S by *Xho*I digestion of a mismatched-PCR 118-bp product and electrophoresis on a 2.5% Nu-Sieve agarose gel. The presence of the mutation creates a *Xho*I restriction site, producing digestion fragments of 17 bp (not shown) and 101 bp. C, healthy control; F, father; M, mother; P, patient; B, brother. (B) Southern blot analysis using a 1.8-kb RT-PCR fragment including all the coding region of the *GBA* gene as a probe. Genomic DNA from a healthy control individual (C) and the patient's mother (M) or the patient (P) was single digested with *Ssp*I, *Bam*HI, and *Xba*I. The lower intensity of a 17.1-kb band (*Ssp*I) and the appearance of extra bands of 3 kb (*Bam*HI) and 3.6 kb (*Xba*I) suggest the presence of a rearrangement between the *GBA* gene and its highly homologous pseudogene.

Mutation Characterization

The patient was originally misdiagnosed as homozygous for the common N370S mutation (12) as the PCR-based analysis showed only the pattern corresponding to the N370S allele. However, family data showed that only the father was a N370S carrier (Fig. 1A), suggesting a deletion in the maternal allele.

To characterize this deletion, Southern analyses with enzymes *Ssp*I, *Xba*I, and *Bam*HI were performed. Note that due to limited material from the patient (P), some of the experiments were

Blood Cells, Molecules, and Diseases (2000) **26**(5) October: 409–416 Cormand et al.
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

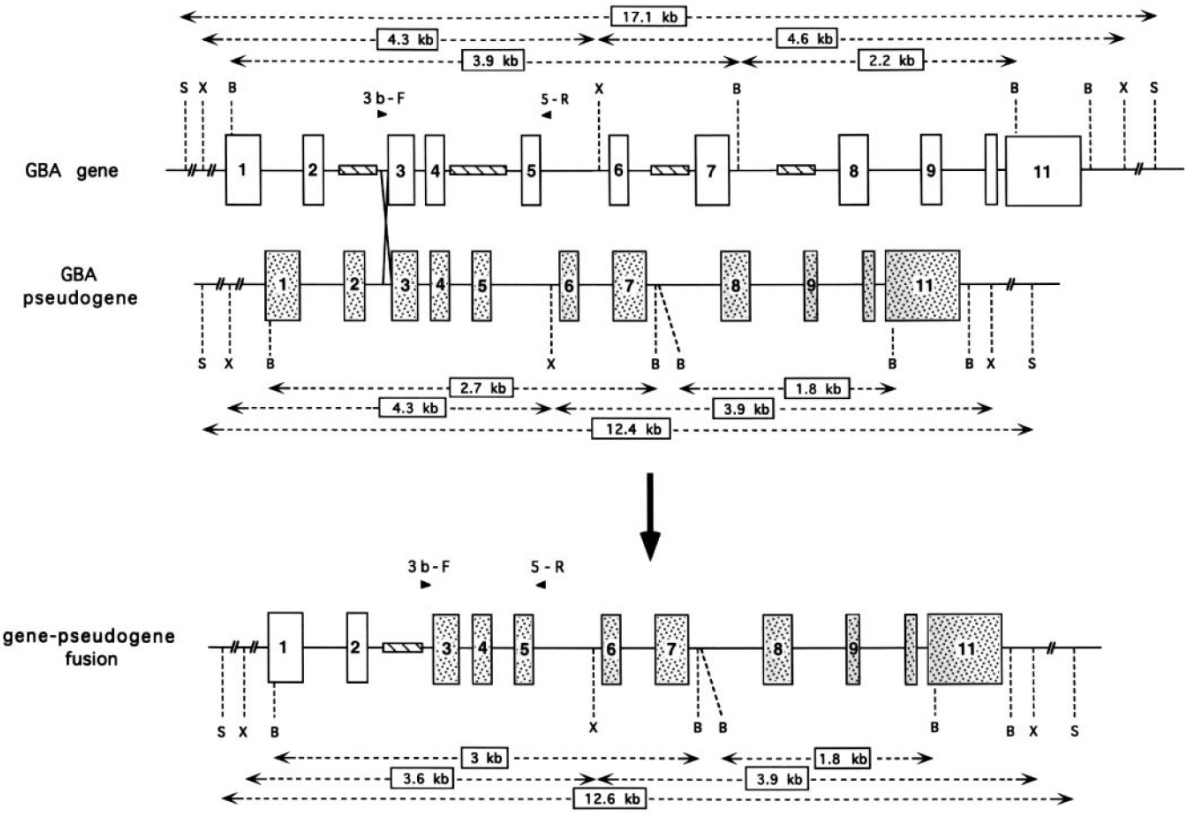


FIG. 2. Schematic representation of the *GBA* gene (open boxes), the *GBA* pseudogene (shadowed boxes) and the gene-pseudogene fusion. The predicted *Ssp*I (S), *Bam*HI (B) and *Xba*I (X) restriction and the expected sizes of the digestion products are shown. *Alu* sequences within the gene are depicted as hatched boxes. The location of PCR primers 3b-F (specific for the gene) and 5-R (nonspecific) used in Fig. 3 and the crossover site are indicated.

performed using maternal DNA (M), which has the same deletion as the patient. *Ssp*I digestion (Fig. 1B) showed a two-band pattern of about 17 (corresponding to the gene) and 12 kb (corresponding to the pseudogene), similar to that of the wild type (Fig. 2). The lower intensity of the upper band suggested the complete deletion of the gene. However, digestion analysis with other enzymes, such as *Bam*HI and *Xba*I, showed different patterns when compared to control DNA. In particular for *Bam*HI, an extra band of approximately 3 kb was obtained, while an extra band of 3.6 kb was found with *Xba*I (Fig. 1B). Moreover, some of the bands, such as the *Xba*I 3.9 kb band in the patient DNA, are more intense than others. These results suggest that a genomic rearrangement, different from a complete deletion of the gene, should have taken place.

The data are consistent with a crossover occurring between the first and the second *Alu* sequences of the gene and its homologous region in the pseudogene. The size of the bands corresponding to the normal and rearranged *GBA* genomic region digested with *Ssp*I, *Bam*HI, and *Xba*I are indicated in Fig. 2. Note that the *Ssp*I fragment containing the fusion gene has a similar, but not identical, size to that of the wild type pseudogene fragment. For the exact length of the fragments, distances between restriction sites were calculated from the published 75-kb sequence of the *GBA* region (GenBank Accession No. AF023268; Ref. 17).

We performed a PCR experiment in order to confirm the hypothesis that the gene-pseudogene crossover occurred in a site located between introns 2 and 4 of the gene. While the forward

Cormand et al. Blood Cells, Molecules, and Diseases (2000) 26(5) October: 409–416
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

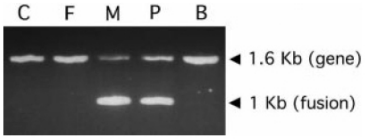


FIG. 3. Characterization of the *GBA* gene-pseudogene fusion by PCR amplification. The location of primers is indicated in Fig. 2. A 1-kb fusion band was obtained from the mother (M) and the patient (P), whereas the father (F) and an unaffected brother (B) only showed the 1.6-kb gene band. A healthy control individual (C) was included.

primer (intron 2) is specific for the gene, the reverse primer (intron 5) is not specific. Amplification on control genomic DNA produces a 1.6-kb fragment from the gene, and no product from the pseudogene. If the hypothesis is correct, a 1.0-kb band should be produced from the fusion gene. Figure 3 shows that both the patient and his mother bear this fusion product, whereas the father and the brother do not.

Upon cloning and sequencing the fusion product, a precise localization of the crossover was obtained. As shown in Fig. 4, the recombinant allele bears an "A" at position 1021 (as in the normal gene sequence) and a dinucleotide "GT," present in the pseudogene, instead of a "CCC" trinucleotide at positions 1040–1042 of the gene. All downstream sequences belong to the pseudogene. These data close down the point of the crossover to nucleotides 1022–1039. The fact that this 18-nt region shows no differences between gene and pseudogene precludes any further refinement.

DISCUSSION

Several complex alleles of the glucocerebrosidase gene have been described. The two most common alleles, Rec*Nci*I and Rec*TL,* are identified according to the pseudogene variants they bear. However, this classification does not take into account neither the actual site of the recombination nor the molecular mechanisms by which these alleles were generated. Although Southern blot analyses or long-PCR amplification could help to understand the mechanism underlying these mutational events, only in a few cases has this issue been addressed (for example, Refs. 5, 9, 18). Most of the reported complex alleles involve a crossover in the 3′ region of the gene, beyond exon 8 (1, 7, 8). As suggested, recombinational hotspots could be present in this region. However, it is also possible that, due to a technical bias, some of the crossovers occurring farther upstream could be missed. Usually, the presence of recombinant alleles is only investigated when L444P, a mutation normally present at the 3′ end of the pseudogene, has been detected. However, as the usual detection protocol for this mutation involves digestion of a PCR-amplified fragment, it could be missed in alleles bearing pseudogene sequences in the annealing site for the gene-specific primers. To our knowledge, the most upstream crossover described so far was located in intron 3 (9). Here we describe a new complex allele which is due to a recombinant event in a more upstream position, in particular between intron 2/exon 3 of the gene and the homologous site of the pseudogene.

The patient was originally diagnosed as an N370S homozygote (12). This misdiagnosis was due to the lack of amplification of the maternal allele. Similar cases have been also described (19, 20). In all these situations, the severe symptoms presented by the patient or the early onset of the disease prompted a more detailed analysis which allowed correction of the original genotype determination. The combination of Southern blot anal-

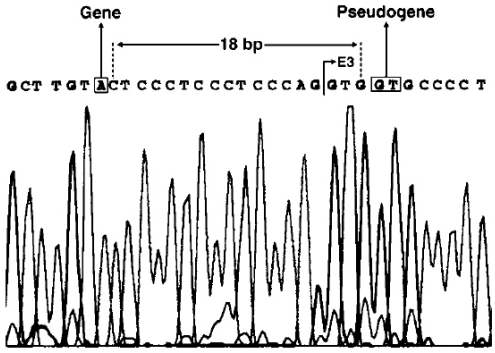


FIG. 4. Determination of the *GBA* gene-pseudogene crossover site by sequence analysis. The 1-kb PCR product shown in Fig. 3 was sequenced using primer 3b-F, which recognizes a specific sequence in intron 2 of the *GBA* gene. The position of the crossover was narrowed down to a genomic interval of 18 bp between a gene-specific nucleotide (A) and pseudogene-specific nucleotides (GT).

Blood Cells, Molecules, and Diseases (2000) **26**(5) October: 409–416
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

TABLE 1

Complex Alleles Reported to Date

First crossover[a]	Second crossover[b]	Exons affected	Allele names	Pseudogene changes in the fusion (or converted) gene[c]	References
1022–1039 (int 2–ex 3)	—	3–11	Rec (int-2)	33 mutations 13 polymorphisms	This study
Intron 3	—	4–11	Rec A	19 mutations 8 polymorphisms	(7, 9)
2456–2476 (int 4–ex 5)	2965–3578 (ex 6–int 6)	5–6	Complex C	R120W, W184R, N188K, V191G, S196P, G202R, F213I	(24)
Intron 6	—	7–11	Complex I	7 mutations 5 polymorphisms	(10)
Intron 8–exon 9	—	9–11	Rec B	del55, D409H, L444P, A456P, V460V	(7)
Intron 8–exon 9	—	9–11	Rec(g4889–6506)	del55, D409H, L444P, A456P, V460V	(25)
5005–5294 (int 8–ex 9)	5690–5722 (int 9)	9	Rec[1263del55, 1342G>C]	del55, D409H, L444P, A456P, V460V	(26)
5005–5294 (int 8–ex 9)	5900–6731 (ex 10–ex 11)	9–10/11	c1263del+Rec*TL*	del55, D409H, L444P, A456P, V460V	(27)
5349–5373 (ex 9)	—	9–11	Rec*TL* Complex B	D409H, L444P, A456P, V460V	(6, 28)
Exon 9	—	9–11	Rec C	D409H, L444P, A456P, V460V	(7)
5374–5689 (ex 9–int 9)	5886–5898 (ex 10)	9–10	RecA456P	D409H, L444P, A456P	(27)
5374–5689 (ex 9–int 9)	—	10–11	Rec*Nci*I Complex A	L444P, A456P, V460V	(5, 6, 28)
Intron 9	—	10–11	Rec D	L444P, A456P, V460V	(7)
5689–5723 (int 9)	—	10–11	Rec allele of patient 1043	L444P, A456P, V460V	(29)
Intron 9	—	10–11	Rec E	L444P, A456P, V460V	(7)
Intron 9–exon 10	—	10–11	Rec F	L444P, A456P, V460V	(7)

Note. int, intron; ex, exon.
[a] Location of first crossover or 5′ limit of gene-converted region.
[b] Location of second crossover or 3′ limit of gene-converted region. (—) Indicates nonexistant or not described crossover.
[c] Only mutations in the coding sequence or splice sites were considered.

ysis, PCR amplification and sequencing allowed the complete characterization of this mutant: it is a recombinant allele with the crossover point located within a 18-bp interval at the end of intron 2 and beginning of exon 3. This allele, which we have named Rec(int-2), is a null allele, as it includes a frameshift mutation early in exon 3 (g.1040delC). All the band sizes observed in Southern analyses are consistent with this crossover site according to the restriction site data within the 75-kb published sequence (17). However, the higher intensity of the 4.3 kb band obtained after control DNA digestion with *Xba*I (see Fig. 1B) suggests the existence of an extra *Xba*I site, 5′ of the pseudogene, not present in the reported sequence. We propose that a sequencing error or a polymorphism could have occurred in the nucleotide 52090 of the 75-kb sequence, converting a C to a G and the sequence "TCTA**C** A" (nt. 52086–52091) to "TCTA**G** A," a *Xba*I site. This would generate a 4248-bp fragment that would comigrate with the 4257 bp fragment of the gene (see Fig. 2).

Large deletions are difficult to characterize. Particularly, it may be complicated to differentiate a fusion gene generated by a crossover in the 5′ region of the gene from a complete deletion of the gene, as a PCR amplification of part of the gene could fail to make this distinction, depending on the primers and the site of the crossover. Moreover, the *Ssp*I analysis, frequently used to detect rearrangements, could also be non-informative, as shown in the case presented in this work (Fig. 1B).

Several complex alleles have been published so far. Most of them are listed in Table 1, not including those reported only in abstracts or con-

Blood Cells, Molecules, and Diseases (2000) **26**(5) October: 409–416
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

gress communications. In some cases, an internal part of the gene has been replaced by the homologous sequence of the pseudogene. Both gene conversion or double crossover events have been suggested as putative mutational mechanisms. The two crossover sites (or the two limits of the converted region) are indicated for them in Table 1. On the other hand, complex alleles presenting gene sequences at their 5′ end and pseudogene sequences at the 3′ end, may derive from a single crossover event. They are consistent with the generation of a gene-pseudogene fusion resulting in the deletion of the chromosomal material between them. Characterization of this kind of allele may be difficult and it has seldom been performed. In some cases the recombination site has not been precisely identified. Thus, some of the complex alleles found in independent patients and bearing the same combination of pseudogene mutations could represent indeed the same allele. In the case presented here we were able to confirm the generation of a fusion gene and to define the crossover region precisely in the 5′ end of the *GBA* gene.

An increasing number of reports suggest that recombinant alleles are a frequent cause of mutation in GD (18, 21–23), and also a possible source of genotyping misdiagnosis (11, 19, 20). The incorrect assignment of mutated alleles may eventually lead to confusing clinical correlations, and further complicate the prognosis of the disease. A focused search and complete characterization of complex alleles, including Southern blot and long PCR methods, would shed light both on mutation mechanisms and genotype-phenotype correlations in Gaucher disease.

ACKNOWLEDGMENTS

We thank Dr. Pedrola, Dr. Pineda, and Dr. Toll for their contribution to the clinical data and Robin Rycroft for revising the English. This work was supported by CICYT (SAF 97-0074), by Fundació Marató de TV3 (Project Nos. 98-1220 and 1221), and by Genzyme, S.L.

REFERENCES

1. Beutler, E., and Gelbart, T. (1998) Hematologically important mutations: Gaucher disease. *Blood Cells Mol. Dis.* **24,** 2–8.
2. Strasberg, P. M., Skomorowski, M. A., Warren, I. B., Hilson, W. L., Callahan, J. W., and Clarke, J. T. R. (1994) Homozygous presence of the crossover (fusion gene) mutation identified in a type II Gaucher disease fetus: Is this analogous to the Gaucher knock-out mouse model? *Biochem. Med. Metab. Biol.* **53,** 16–21.
3. Sidransky, E., Tayebi, N., Stubblefield, B. K., Eliason, W., Klineburgess, A., Pizzolato, G. P., Cox, J. N., Porta, J., Bottani, A., and Delozierblanchet, C. D. (1996) The clinical, molecular, and pathological characterisation of a family with two cases of lethal perinatal type 2 Gaucher disease. *J. Med. Genet.* **33,** 132–136.
4. Finn, L. S., Zhang, M., Chen, S. H., and Scott, C. R. (2000) Severe Type II Gaucher disease with ichthyosis, arthrogryposis and neuronal apoptosis: Molecular and pathological analyses. *Am. J. Med. Genet.* **91,** 222–226.
5. Zimran, A., Sorge, J., Gross, E., Kubitz, M., West, C., and Beutler, E. (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.
6. Latham, T., Grabowski, A., Theophilus, B. D. M., and Smith, F. I. (1990) Complex alleles of the acid β-glucosidase gene in Gaucher disease. *Am. J. Hum. Genet.* **47,** 79–86.
7. Tayebi, N., Stone, D. L., and Sidransky, E. (1999) Type 2 Gaucher disease: An expanding phenotype. *Mol. Genet. Metab.* **68,** 209–219.
8. Stone, D. L., Tayebi, N., Orvisky, E., Stubblefield, B., Madike, V., and Sidransky, E. (2000) Glucocerebrosidase gene mutations in patients with type 2 Gaucher disease. *Hum. Mutat.* **15,** 181–188.
9. Reissner, K., Tayebi, N., Stubblefield, B. K., Koprivica, V., Blitzer, M., Holleran, W., Cowan, T., Almashanu, S., Maddalena, A., Karson, E. M., and Sidransky, E. (1998) Type 2 Gaucher disease with hydrops fetalis in an Ashkenazi Jewish family resulting from a novel recombinant allele and a rare splice junction mutation in the glucocerebrosidase. *Mol. Genet. Metab.* **63,** 281–288.
10. Filocamo, M., Bonuccelli, G., Mazzotti, R., Giona, F., Gatti, R. Identification of a novel recombinant allele in three unrelated Italian Gaucher patients: implications for prognosis and genetic counseling. *Blood Cells Mol. Dis.,* in press.
11. Cormand, B., Grinberg, D., Gort, L., Chabás, A., and Vilageliu, L. (1998) Molecular analysis and clinical findings in the Spanish Gaucher disease population: Putative haplotype of the N370S ancestral chromosome. *Hum. Mutat.* **11,** 295–305.
12. Cormand, B., Vilageliu, L., Burguera, J. M., Balcells, S., González-Duarte, R., Grinberg, D., and Chabás, A.

Blood Cells, Molecules, and Diseases (2000) **26**(5) October: 409–416 Cormand et al.
doi:10.1006/bcmd.2000.0317, available online at http://www.idealibrary.com on IDEAL®

(1995) Gaucher disease in Spanish patients: Analysis of eight mutations. *Hum. Mutat.* **5,** 303–309.

13. Miller, S. A., Dyke, D. D., and Polesky, H. F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16,** 1215.
14. Horowitz, M., Wilder, S., Horowitz, Z., Reiner, O., Gelbart, T., and Beutler, E. (1989) The human glucocerebrosidase gene and pseudogene: Structure and evolution. *Genomics* **4,** 87–96.
15. Beutler, E., West, C., and Gelbart, T. (1992) Polymorphisms in the human glucocerebrosidase gene. *Genomics* **12,** 795–800.
16. Beutler, E., Gelbart, T., and West, C. (1990) The facile detection of the nt 1226 mutation of glucocerebrosidase by "mismatched" PCR. *Clin. Chim. Acta* **194,** 161–166.
17. Winfield, S. L., Tayebi, N., Martin, B. M., Ginns, E. I., and Sidransky, E. (1997) Identification of three additional genes contiguous to the glucocerebrosidase locus on chromosome 1q21: Implications for Gaucher disease. *Genome Res.* **7,** 1020–1026.
18. Cormand, B., Harboe, T. L., Gort, L., Campoy, C., Blanco, M., Chamoles, N., Chabás, A., Vilageliu, L., and Grinberg, D. (1998) Mutation analysis of Gaucher disease patients from Argentina: High prevalence of the RecNciI mutation. *Am. J. Med. Genet.* **80,** 343–351.
19. Beutler, E., and Gelbart, T. (1994) Erroneous assignment of Gaucher disease genotype as a consequence of a complete gene deletion. *Hum. Mutat.* **4,** 212–216.
20. Lecoutre, P., Demina, A., Beutler, E., Beck, M., and Petrides, P. (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.
21. Horowitz, M., and Zimran, A. (1994) Mutations causing Gaucher disease. *Hum. Mut.* **3,** 1–11.
22. Nelson, P. V., Carey, W. F., Morris, C. P., and Lewis, B. D. (1995) Mutation analysis of Australasian Gaucher disease patients. *Am. J. Med. Genet.* **58,** 382.
23. Koprivica, V., Stone, D. L., Park, J. K., Callahan, M., Frisch, A., Cohen, I. J., Tayebi, N., and Sidransky, E. (2000) Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease. *Am. J. Hum. Genet.* **66,** 1777–1786.
24. Latham, T. E., Theophilus, B. D. M., Grabowski, G. A., and Smith, F. I. (1991) Heterogeneity of mutations in the acid β-glucosidase gene of Gaucher disease patients. *DNA Cell Biol.* **10,** 15–21.
25. Hodanova, K., Hrebicek, M., Cervenkova, M., Mrazova, L., Veprekova, L., and Zeman, J. (1999) Analysis of the β-glucocerebrosidase gene in Czech and Slovak Gaucher patients: Mutation profile and description of six novel mutant alleles. *Blood Cells Mol. Dis.* **25,** 287–298.
26. Sarria, A. J., Giraldo, P., Perez-Calvo, J. I., and Pocoví, M. (1999) Detection of three rare (G377S, T134P and1451delAC), and two novel mutations (G195W and Rec[1263del55;1342G>C] in Spanish Gaucher disease patients. *Hum. Mutat.* **14,** 88 (Online: Mutation in Brief #251).
27. Hatton, C. E., Cooper, A., Whitehouse, C., and Wraith, J. E. (1997) Mutation analysis in 46 British and Irish patients with Gaucher's disease. *Arch. Dis. Child* **77,** 17–22.
28. Eyal, N., Wilder, S., and Horowitz, M. (1990) Prevalent and rare mutations among Gaucher patients. *Gene* **96,** 277–283.
29. Sinclair, G., Choy, F., and Humphries, L. (1998) A novel complex allele and two new point mutations in type 2 (acute neuronopathic) Gaucher disease. *Blood Cells Mol. Dis.* **24,** 420–427.

1.2. Caracterització del mecanisme de generació de l'al·lel Rec*Nci*I en població argentina i espanyola

Una de les mutacions causants de la malaltia de Gaucher és l'al·lel recombinant Rec*Nci*I. Aquest al·lel és un al·lel complex que es defineix per contenir 3 canvis (L444P, A456P i V460V). Aquests tres canvis estan presents en el pseudogèn *GBA* (*GBAP*). Hi ha diferents mecanismes pels quals aquests canvis passen al gen i esdevenen patogèniques. Bàsicament aquests mecanismes són la conversió gènica i la recombinació o entrecreuament desigual. Depenent de quin mecanisme el generi, l'estructura de l'al·lel és diferent, i es pot posar de manifest pels diferents patrons de *Southern blot* generats amb l'enzim *Ssp*I.

Ens vam proposar analitzar els mecanismes que generaven aquests diferents patrons dels al·lels Recs en els pacients portadors de l'al·lel Rec*Nci*I. Disposàvem de mostres de 25 pacients d'origen argentí, ja que en aquesta població la freqüència d'aquesta mutació és especialment elevada (Cormand i col., 1998b), i també mostres de 3 pacients de la població espanyola portadors de la mutació Rec*Nci*I.

També vam analitzar mostres de pacients portadors de la mutació L444P, tant de la població argentina com de la població espanyola. En total vam analitzar 34 pacients argentins i 27 espanyols, entre portadors de l'al·lel Rec*Nci*I i la mutació L444P.

Es va posar de manifest que la majoria d'al·lels Rec*Nci*I s'havien produït per conversió gènica i que en alguns dels al·lels analitzats, Rec*Nci*I o L444P, s'havien produït reordenaments que implicaven el gen de la metaxina.

Referència del treball publicat:

Gene rearrangements in the glucocerebrosidase-metaxin region giving rise to disease-causing mutations and polymorphisms. Analysis of 25 Rec*Nci*I alleles in Gaucher disease patients. *Human Genetics (2003) 112: 426-429*.

Hum Genet (2003) 112 : 426–429
DOI 10.1007/s00439-002-0894-0

SHORT REPORT

**Anna Díaz-Font · Bru Cormand · Mariana Blanco
Néstor Chamoles · Amparo Chabás · Daniel Grinberg
Lluïsa Vilageliu**

Gene rearrangements in the glucocerebrosidase-metaxin region giving rise to disease-causing mutations and polymorphisms. Analysis of 25 Rec*Nci*I alleles in Gaucher disease patients

Received: 8 August 2002 / Accepted: 15 November 2002 / Published online: 14 February 2003

Abstract The glucocerebrosidase and metaxin genes lie in a gene-rich region that also includes two corresponding pseudogenes. This gives rise to recombinant alleles. We analysed two groups of patients from Argentina and Spain: 25 bearing the Rec*Nci*I allele and 36 carrying L444P. The mutational mechanism is described and the crossover site precisely defined. Most of the Rec*Nci*I alleles were generated by gene conversion. Rearranged alleles involving the metaxin gene were also identified. The high frequency of Rec*Nci*I alleles associated with a polymorphic rearrangement at the metaxin level is probably due to a founder effect.

Introduction

The most prevalent glucocerebrosidase (*GBA)* complex allele causing Gaucher disease (GD) is Rec*Nci*I (Zimran et al. 1990). Although it is not frequent in most populations, it has been found at a high frequency (21%) among Argentinian GD alleles (Cormand et al. 1998b). This allele is generated either by gene conversion or by unequal crossing over between the *GBA* gene and the pseudogene (*GBAP*) (Zimran et al. 1990). Southern blot and microsatellite analyses may identify which mechanism is involved. While gene conversion produces a wild-type pattern, a *GBA-GBAP* unequal crossover generates a pattern that corresponds to that of a fusion gene. Additionally, a third pattern, the result of a crossover between the metaxin and the pseudometaxin genes, has been reported. It has been suggested that this is a non-Gaucher disease-causing variant (Tayebi et al. 2000).

A. Díaz-Font · B. Cormand · D. Grinberg · L. Vilageliu (✉)
Departament de Genètica, Facultat de Biologia,
Universitat de Barcelona,
Av. Diagonal, 645, E-08028 Barcelona, Spain
Tel.: +34-93-4021494, Fax: +34-93-4110969,
e-mail: lvilageliu@ub.edu

M. Blanco · N. Chamoles
Fundación para el estudio de las Enfermedades Neurometabólicas,
Uriarte 2379, 1425 Buenos Aires, Argentina

A. Chabás
Institut de Bioquímica Clínica, Hospital Clínic,
Corporació Sanitària, Mejia Lequerica s/n, Edificio Helios III,
E-08028 Barcelona, Spain

Subjects and methods

Thirty-four Argentinian (22 bearing the Rec*Nci*I allele and 12 bearing the L444P mutation) and 27 Spanish (three carrying Rec*Nci*I and 24 L444P) GD patients were analysed for the presence of genomic rearrangements. Some have been described elsewhere (Cormand et al. 1998a, 1998b).

DNA preparation, mutation detection, PCR amplification, sequencing, and microsatellite and Southern blot analyses were performed as described (Cormand et al. 1998a, 1998b; Rodriguez-Marí et al. 2001). Sequence numbering: GenBank AF023268.

Gene mutation nomenclature used in this article follows the recommendations of den Dunnen and Antonarakis (2001). Gene symbols used in this article follow the recommendations of the HUGO Gene Nomenclature Committee (Povey et al. 2001).

Results and discussion

In order to characterise the mechanisms generating the Rec*Nci*I alleles, DNA samples were analysed by Southern blot using the *Ssp*I restriction enzyme and/or by determining the number of alleles at the ITG6.2 polymorphism described by Lau et al. (1999).

The data obtained from the 25 GD patients carrying the Rec*Nci*I allele showed that nine of them presented either Southern and/or ITG6.2 patterns similar to those of wild-type individuals, suggesting that the Rec*Nci*I allele was generated by gene conversion (structure A in Fig. 1A). Fifteen presented either the 17.1/15.6/12.4-kb Southern pattern and/or the three-allele ITG6.2 combination, indicating the presence of a duplication (structure C in Fig. 1A, B). Only one Rec*Nci*I allele (patient A48) appeared to have been caused by a recombination event leading to a fusion gene (structure B1). This was suggested by the 13.9-kb Southern band and the lack of one ITG6.2 allele, and was further confirmed by a specific 4.6-kb PCR amplification.

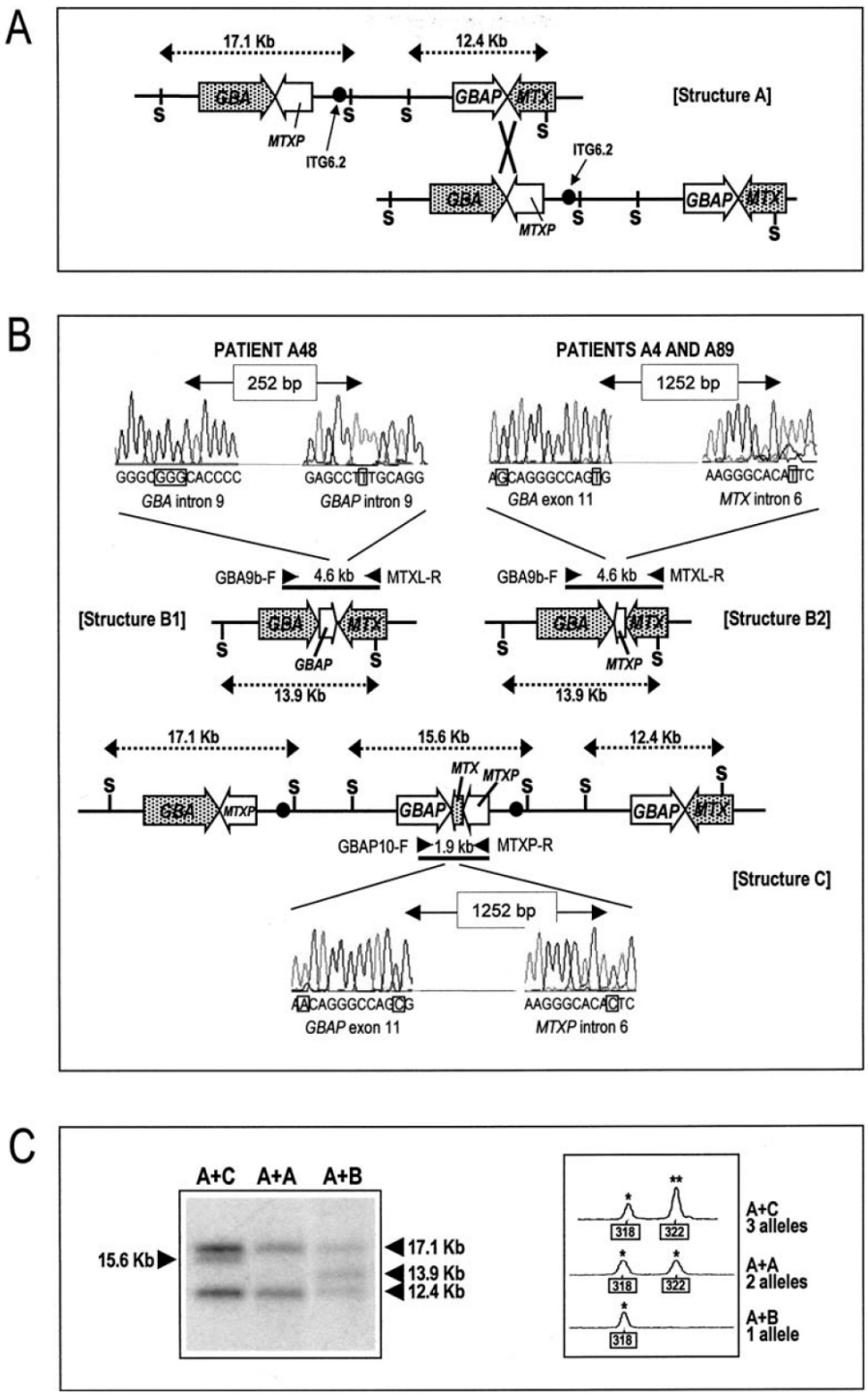


Fig. 1 **A** Genomic region including the *GBA* and *MTX* genes and pseudogenes, showing the unequal crossover between the *GBA-MTX* clusters that gives rise to rearranged structures. *S* indicates *Ssp*I restriction sites. Lengths of restriction fragments are indicated by *dashed lines*. Genes are depicted in *grey* and pseudogenes in *white*. Transcription direction is indicated. *Structure A* corresponds to that found in wild-type chromosomes. A gene conversion event between the *GBA* pseudogene and gene would not alter this genomic structure. **B** Different genomic structures (*B1*, *B2* and *C*) in rearranged chromosomes. Genome structures B1 and B2 generate a 13.9-kb band when analysed by Southern blot. This rearrangement implies a deletion of the whole region between the two gene clusters, eliminating the ITG6.2 marker. Structure C (the reciprocal product of B2) produces a three-band Southern pattern with an additional 15.6-kb band. The sequences correspond to gene/pseudogene boundaries within the amplified fragments. Each chromatogram includes positions (*boxed*) that are different in the gene and pseudogene. Minimal intervals for crossover sites are indicated. Genotypes of patients bearing structures B1 and B2 are A48: N370S/Rec*Nci*I, A4: N370S/L444P and A89: G377S/L444P. The specific 4.6-kb and 1.9-kb PCR fragments and primers are also indicated. Primer sequences and PCR conditions are available on request. **C** *Left* Southern blot from individuals presenting structure A in one chromosome and either structure A (patient A57, N370S/Rec*Nci*I), B (patient A89) or C (patient A77, N370S/Rec*Nci*I) in the other. *Right* Chromatogram showing ITG6.2 analysis for the same individuals. *Asterisks* on ITG6.2 alleles correspond to number of alleles per peak. Allele sizes are shown *below the peaks*. For A89 (A+B), parental analysis (not shown) indicated that the chromatogram corresponded to only one allele

The genomic structure was also analysed in samples from the 36 patients homozygous or heterozygous for the L444P mutation. Most presented the same structure as the wild-type samples. Surprisingly, different results were found in three samples. Two of these (A4 and A89) presented the 17.1/13.9/12.4-kb Southern pattern, only one ITG6.2 allele and the 4.6-kb PCR amplification fragment indicative of a B rearrangement. As they bore only the L444P mutation and not the two additional changes of the Rec*Nci*I alleles, the crossover must have taken place be-

Table 1 Genotypes for *GBA* mutations and markers ITG6.2, 5GC3.2 and D1S1595 for individuals bearing the duplication. Individuals without the Rec*Nci*I mutation are listed separately. Alleles shown in *bold* are assumed to be in phase with the duplication (according to parental data or N370S association analysis, i.e. N370S with 318 and 222 as reported in Lau et al. 1999 and Rodriguez-Mari et al. 2001). For A1 and A80, the phase could not be established for one allele (*N.D.* = not determined)

Individuals	Type	Mutations	ITG6.2	5GC3.2	D1S1595
A1	2	Rec*Nci*I/D409H	318 **322** 322	**222** 222	7 **8**
A12	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	5 **8**
A15	1	Rec*Nci*I/N370S	**322 322** 322	**220** 222	5 **7**
A16	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	3 8
A28	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	5 **8**
A44	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	5 **8**
A55	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	5 **8**
A69	1	Rec*Nci*I/N370S	**322 322** 322	**222** 222	5 **8**
A70	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	5 **8**
A71	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	5 **8**
A74	1	Rec*Nci*I/N370S	**322 322** 322	N.D.	N.D.
A77	1	Rec*Nci*I/N370S	**322 322** 322	**222** 222	3 7
A80	1/3[a]	Rec*Nci*I/F411I	318 **322** 322	**222** 222	5 **7**
A85	1	Rec*Nci*I/N370S	*318* **322 322**	**222** 222	3 **8**
A109	1	Rec*Nci*I/N370S	*318* **322 322**	N.D.	5 **8**
I.37	1	L444P/N370S	*318* **322 322**	**222** 222	**5** 5
Contr.19G	–	+/+	318 **318** 322	**222** 222	2 5

[a]No neurological symptoms at three years of age

yond the 3′ end of the *GBA* gene, giving rise to structure B2. This rearrangement is expected to be non-pathogenic and independent of the L444P mutation. The third sample (I.37) presented three ITG6.2 alleles, indicating that it bears a duplicated segment (structure C).

The crossover sites for alleles bearing structures B1, B2 and C were narrowed down to the shortest possible fragments. In patient A48, the crossover took place somewhere within a 252-bp fragment in the *GBA*/*GBAP* intron 9, generating structure B1. As shown, at position IVS9+17 there are three Gs, as in the gene (four Gs in the pseudogene), and at IVS9-101 there is a T, as in the pseudogene (C in the gene). This crossover generates the B1 structure shown in Fig. 1B: exons 1–9 of the *GBA* gene fused to exons 10 and 11 of the *GBA* pseudogene (exon 10 is the one where the three Rec*Nci*I changes are located). For patients A4 and A89, the crossover site was located beyond the *GBA* sequence: positions 38,481 (G) and 38,491 (T) correspond to gene sequences (A and C in the *GBAP*). But at position 39,744, within intron 6 of the pseudometaxin gene (or 60,356 of the metaxin gene), they bore the metaxin sequence (T), confirmed by *Dra*III digestion (data not shown). Thus, the crossover site was narrowed down to a 1,252-bp interval, generating structure B2. For alleles bearing structure C, a specific PCR amplification was performed to yield a 1.9-kb fragment, which was analysed by *Hha*I and *Dra*III digestion (not shown) and sequencing. The crossover site was within the same sequence interval as that for structure B2, but in this case the reciprocal product was generated. Positions 38,481 (A) and 38,491 (C) in exon 11 correspond to *GBAP,* while position 39,744 (C) corresponds to the pseudometaxin gene. This rearrangement is presumably not pathogenic since a copy of the *GBA* gene remains intact. Thus, the presence of the disease-causing mutations in the patients must have arisen independently.

As structures B2 and C are not directly associated with Gaucher disease they may be found in the control population. Yet, when the presence of these structures was analysed in 49 healthy Spanish individuals, no B2 or C alleles were found. However, as all 15 patients bearing the C structure and the Rec*Nci*I allele were Argentinian with Spanish surnames, a control population from Galicia (the origin of most Spanish immigrants to Argentina) was also analysed. In this sample we found one structure-C allele (and also one structure-B) among 86 chromosomes.

After finding that more than 50% of the Rec*Nci*I alleles bore the duplication, haplotype analysis was performed in order to examine whether the high frequency of this complex, rearranged allele could be explained by a founder effect. Markers ITG6.2, 5GC3.2 (Lau et al. 1999), and D1S1595 were used. Analysis of "Rec*Nci*I + duplication (structure C)" alleles (Table 1) showed that ten out of 13 bore the same haplotype: 322-322-222, for the duplicated ITG6.2 and for the 5GC3.2 markers, respectively, and genotypes of two other patients are consistent with this haplotype. For marker D1S1595, ten out of 14 bear allele 8 in phase with mutation Rec*Nci*I. The alleles 322 (ITG6.2), 222 (5GC3.2) and 8 (D1S1595) are over-represented in the "Rec*Nci*I + duplication" alleles. This association proved significant in a chi-square test for the 322 (P=0.002), and 8 (P<0.001) alleles, but not for the 222 (data not shown). These data suggest a single origin for the duplicated alleles associated with a Rec*Nci*I mutation. The duplication was also found together with mutation L444P in one patient (I.37) and in one control individual (19G).

In summary: we found that most Rec*Nci*I alleles were generated by gene conversion, we mapped the precise crossover site on the rearranged alleles and we found that a founder effect might account for the high frequency of Rec*Nci*I alleles associated with the duplication at the metaxin level.

Acknowledgements The authors are grateful to Dr. A. Carracedo and D. Gallardo for providing control samples, to L. Gort for technical assistance and to R. Rycroft for revising the English. They

also thank Serveis Científico-Tècnics, Universitat de Barcelona. ADF is a recipient of a fellowship from the Spanish government. This research was supported by CICYT (SAF 2000-0200), Fundació Marató de TV3 (Proj. Nos. 98-1220/1221), and Genzyme. All experiments comply with current Spanish legislation.

References

Cormand B, Grinberg D, Gort L, Chabás A, Vilageliu L (1998a) Molecular analysis and clinical findings in the Spanish Gaucher disease population: putative haplotype of the N370S ancestral chromosome. Hum Mutat 11:295–305

Cormand B, Harboe TL, Gort L, Campoy C, Blanco M, Chamoles N, Chabás A, Vilageliu L, Grinberg D (1998b) Mutation analysis of Gaucher disease patients from Argentina: High prevalence of the RecNciI mutation. Am J Med Genet 80:343–351

den Dunnen JT, Antonarakis SE (2001) Nomenclature for the description of human sequence variations. Hum Genet 109:121–124

Lau EK, Tayebi N, Ingraham LJ, Winfield SL, Koprivica V, Stone DL, Zimran A, Ginns EI, Sidransky E (1999) Two novel polymorphic sequences in the glucocerebrosidase gene region enhance mutational screening and founder effect studies of patients with Gaucher disease. Hum Genet 104:293–300

Povey S, Lovering R, Bruford E, Wright M, Lush M, Wain H (2001) The HUGO Gene Nomenclature Committee (HGNC). Hum Genet 109:678–680

Rodríguez-Marí A, Díaz-Font A, Chabás A, Pastores GM, Grinberg D, Vilageliu L (2001) New insights into the origin of the Gaucher disease-causing mutation N370S: Extended haplotype analysis using the 5GC3.2, 5470 G/A, and ITG6.2 polymorphisms. Blood Cells Mol Dis 27:950–959

Tayebi N, Park J, Madike V, Sidransky E (2000) Gene rearrangement on 1q21 introducing a duplication of the glucocerebrosidase pseudogene and a metaxin fusion gene. Hum Genet 107: 400–403

Zimran A, Sorge J, Gross E, Kubitz M, West C, Beutler E (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