WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

An evolutionary and structure-based docking model for glucocerebrosidase–saposin C and glucocerebrosidase–substrate interactions—Relevance for Gaucher disease

Sílvia Atrian,[1,2*] Eduardo López-Viñas,[3] Paulino Gómez-Puertas,[3] Amparo Chabás,[4,5] Lluïsa Vilageliu,[1,2,5] and Daniel Grinberg[1,2,5]

[1] Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain

[2] Institut de Biomedicina de la Universitat de Barcelona (IBUB), 08028 Barcelona, Spain

[3] Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Cantoblanco, 28049 Madrid, Spain

[4] Institut de Bioquímica Clínica, Hospital Clinic, Corporació Sanitària Clínic, 08028 Barcelona, Spain

[5] CIBERER, Instituto de Salud Carlos III, Barcelona, Spain

ABSTRACT

Gaucher disease, the most prevalent lysosomal storage disorder, is principally caused by malfunction of the lysosomal enzyme glucocerebrosidase (GBA), a 497-amino acid membrane glycoprotein that catalyzes the hydrolysis of glucosylceramide to ceramide and glucose in the presence of an essential 84-residue activator peptide named saposin C (SapC). Knowledge of the GBA structure, a typical $(\beta/\alpha)_8$ TIM barrel, explains the effect of few mutations, directly affecting or located near the catalytic site. To identify new regions crucial for proper GBA functionality, we analyzed the interactions of the enzyme with a second (substrate) and a third (cofactor) partner. We build 3D docking models of the GBA–SapC and the GBA–ceramide interactions, by means of methodologies that integrate both evolutive and structural information. The GBA–SapC docking model confirm the implication of three spatially closed regions of the GBA surface (TIM barrel-helix 6 and helix 7, and the Ig-like domain) in binding the SapC molecule. This model provides new basis to understand the pathogenicity of several mutations, such as the prevalent Leu444Pro, and the additive effect of Glu326Lys in the double mutant Glu326Lys-Leu444Pro. Overall, 39 positions in which amino acid changes are known to cause Gaucher disease were localized in the GBA regions identified in this work. Our model is discussed in relation to the phenotype (pathogenic effect) of these mutations, as well as to the enzymatic activity of the recombinant proteins when available. Both data fully correlates with the proposed model, which will provide a new tool to better understand Gaucher disease and to design new therapy strategies.

Proteins 2008; 70:882–891.

Key words: correlated mutations; **in silico** *modeling; structure–function relationship; GBA.*

INTRODUCTION

Gaucher disease (GD, MIM# 230800, 230900, 231000), the most prevalent lysosomal storage disorder worldwide, is principally due to a deficiency of the lysosomal enzyme glucocerebrosidase (D-glucosyl acylsphingosine glucohydrolase, EC 3.2.1.45). This enzyme is a 497-amino acid-long membrane glycoprotein of 65 kDa that catalyzes the hydrolysis of glucosylceramide (GlcCer) to ceramide and glucose in the presence of an activator protein named saposin C (SapC). More than 200 mutations have been identified in the *GBA* gene located on 1q21 (http://www.hgmd.org).[1] According to the severity of their phenotypic effect, the mutations have been classified as mild, severe, or lethal.[2] The disease has classically been divided into three types based on neurological involvement: Type 1 (non-neuronopathic), Type 2 (acute neuronopathic), and Type 3 (subacute neuronopathic) (for a review on Gaucher disease, see Beutler and Grabowski[3]). Few genotype–phenotype correlations have been established, such as the prevalent N370S mutation with Type 1, or the L444P allele with the neuronopathic forms of the disease.[4] In addition, mutation D409H in homozygosity has been

Grant sponsor: Ministerio de Educación y Ciencia; Grant numbers: BIO2006-14420-CO2-01, SAF2004-06843, SAF2003-00386, SAF2006-12276; Grant sponsor: Ministerio de Sanidad y Consumo; Grant numbers: PI042350, PI051343, PI040867, PI051182; Grant sponsor: Fundación Ramón Areces.
Sílvia Atrian and Eduardo López-Viñas contributed equally to this work.
Lluïsa Vilageliu and Daniel Grinberg are the co-last authors.
*Correspondence to: Sílvia Atrian, Departament de Genètica, Facultat de Biología, Universitat de Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain.
E-mail: satrian@ub.edu
Received 7 December 2006; Revised 19 March 2007; Accepted 28 March 2007
Published online 5 September 2007 in Wiley InterScience (www.interscience.wiley.com)
DOI: 10.1002/prot.21554

associated with a special Type 3 phenotype presenting severe cardiac involvement and oculomotor apraxia.[5,6] *In vitro* expression analyses of several GBA mutations have been performed to elucidate the effect of the mutation on enzyme activity and to explore phenotype–genotype correlations.[7]

A different and complementary approach to gain insight on the effects of point mutations is to determine the modifications they might cause on the 3D structure of the enzyme. Most of the *GBA* gene defects causing Gaucher disease are missense substitutions.[8] For such mutations, analysis of the alterations they might cause on the enzyme structure is appropriate. The GBA 3D structure was solved in 2003[9] by 2.0 Å X-ray diffraction crystal analysis, revealing a typical $(\beta/\alpha)_8$ TIM barrel catalytic core, common to most members of the GH-A group of lysosomal glycoside hydrolases (Glyco_-hydro_30 Pfam Family; http://www.sanger.ac.uk/Software/Pfam/). This structure encompasses three folding domains: Domain I, a three-strand antiparallel β-sheet flanked by a loop and a perpendicular strand; Domain II, an Ig-like fold formed by two β-sheets; and finally, Domain III, the central $(\beta/\alpha)_8$ TIM barrel. No protein function has been attributed to the first two domains. Knowledge of the GBA structure opened the way toward understanding the molecular basis of the enzyme malfunction underlying lysosomal disorders. However, the mutations that directly affect the catalytic site, or those that are located near the catalytic pocket, constitute a minority. Moreover, there is no evidence or even a hypothesis explaining the deleterious effects of several of the amino acid changes in Domain III, or of all those involving Domains I and II.

Analysis of the interactions of the GBA enzyme with a second (substrate molecule) and third (cofactors) partner will shed light on protein regions important for proper enzymatic function, even for those located far from catalytic residues. GBA activity depends on the presence of saposin C as an enzyme activator (reviewed in Beutler and Grabowski[3]). Saposin C, an 84-residue protein essential for glucosylceramide hydrolysis, exhibits a pH-dependent interaction with the phospholipids vesicles through reversible membrane binding.[10] The mechanistic model of interaction with GBA proposed thus far[11] assumes that an aggregate formed by the enzyme and saposin C molecules would interact with the vesicular surface containing the lipidic molecules, thereby facilitating the interaction of the enzyme and its substrate. However, despite the fact that the saposin C 3D structure has already been solved by NMR,[12] no attempt has been made to structurally characterize the interaction between the GBA and SapC molecules. It is sensible to assume that amino acid substitutions in those residues present in the GBA/SapC interacting surface would impair enzymatic activity and thus determine GD pathology, even if the catalytic efficiency remained unaffected. In fact, Salvioli *et al.*[13] recently observed that N370S, the most prevalent GD mutation, affected the capacity of the enzyme to interact with Sap C and the phospholipid-containing membranes.

In the absence of experimental data for protein–protein interactions at the structural level, *in silico* methods could provide feasible models to explain observed phenotypic characteristics. Despite the lower quality of theoretical models for protein dimerization compared with cocrystallized structures, recent advances in docking methodologies has greatly improved the reliability of *in silico* approaches, as proved in recent CAPRI competitions.[14] This encouraged us to build a 3D docking model of the GBA–Sap C interaction. Our results, based on a integrative method of both evolutive and structural information,[15] identified a patch of residues exhibiting high values in terms of correlated (or concerted) mutations in the SapC and GBA complementary surfaces. Finally, docking models of both the open[12] and closed[16] SapC structures on the GBA surface[9] confirm the implication of three spatially close regions (TIM barrel-helix 6 and helix 7, and the Ig-like domain) in binding the SapC molecule. Furthermore, we used flexible docking strategies to model the substrate/enzyme interactions, thereby permitting us to identify surface regions/residues involved in the correct positioning of the cerebroside molecule, and not simply those involved in the catalytic site.

METHODS

Data sets

Amino acid sequences of acid-β-glucosidase (GBA) and saposin C (Sap C) proteins were obtained from the UniProt Knowledgebase (EBI-EMBL). The 3D solution structures for closed and the open saposin C were obtained from the Protein Data Bank (codes 1M12 and 1SN6, respectively), as were the GBA coordinates (1OGS). Data corresponding to the phenotype caused by several mutations were retrieved from literature as indicated in Tables I and II. Available data on *in vitro* enzyme activity of the corresponding mutant constructs were also included.

Structural model for the molecular interaction between GBA and saposin C

Models for interaction of GBA (1OGS in Protein Data Bank) to both saposin C structures were built using the protein–protein rigid docking method implemented in the Hex program.[17] To reduce the translational–rotational search problem, the initial positioning of the two structures was calculated taking into account the presence of correlated mutations. These were deduced from multiple sequence alignments for the two proteins present in

Table I
Amino Acid Changes Identified as Gaucher-Disease-Causing Mutations Mapped in the sapC-GBA Interacting Surface, According to Our Docking Model

Residue position	Aa replacement (wt → mutant)	Severity	Enzyme activity	References[a]
Region I [Residues 315–326]				
315	Asp → His	Unknown		Ref. 27
318	Ala → Asp	Unknown		Ref. 27
319	Pro → Ala	Unknown		Ref. 8
323	Thr → Ile	Unknown		Ref. 28
324	Leu → Pro	Unknown		Ref. 29
325	Gly → Trp	Unknown	13.9%	Ref. 30
	Gly → Arg	Severe		Ref. 31
326	Glu → Lys	Modifier variant	42%	Refs. 32 and 7
Region II [Residues 365–373 and 399]				
366	Ser → Gly	Mild		Ref. 33
	Ser → Asn	Unknown		Ref. 34
369	Thr → Met	Unknown		Ref. 35
370	Asn → Ser	Mild	4.5–23.4%	Refs. 36 and 7 (and references therein)
371	Leu → Val	Mild		Ref. 37
399	Asp → Asn	Severe		Ref. 38
	Asp → Tyr	Unknown		Ref. 39
Region III [Residues 438–466 and 487]				
444	Leu → Arg	Severe		Ref. 40
	Leu → Pro	Severe	2–18.1%	Refs. 41 and 7 (and references therein)
446	Ala → Pro	Very mild		Ref. 29
451	His → Arg	Unknown		Ref. 42
460	Val → Met	Unknown		Ref. 8
461	Leu → Pro	Unknown		Ref. 8
463	Arg → Cys	Severe	24.5%	Refs. 43 and 30
	Arg → Pro	Unknown		Ref. 8
465	ΔSer	Unknown	5.5%	Ref. 44

[a]When two references are provided for a mutation, the first describes its report as a disease-causing mutation, and the second the expression and enzyme activity studies of the recombinant protein.

eight different species (human, chimpanzee, orangutan, dog, mouse, *Drosophila*, mosquito, and *C. elegans*), essentially as previously described.[15] This procedure guarantees that the Hex filtering algorithm takes into account the spatial arrangement previously selected by the correlated mutation-based method, a fact that otherwise would not be initially considered in a pure shape and electrostatic docking approach. Correlation coefficients of mutations between all pairs of positions in the alignments were calculated using the PLOTCORR program.[18] The generation of 5×10^3 shape-based alternative docking solutions for the dimerization model was achieved using the low-resolution docking algorithm GRAMM.[19] The harmonic average factor (X_d)[20] was calculated for each solution. This factor estimates the spatial proximity of residues, taking into account the distance of both the correlated pairs and all pairs of positions in the alignment. Distances between pairs of residues were grouped in bins of 4 Å for each of the solutions, obtaining two different distributions of binned data for the correlated pairs and for all pairs of positions. The difference between the two distributions was calculated bin-by-bin and normalized to increase the weight of closer distances. The X_d factor for each docking solution was calculated by the formula:

$$X_d = \sum_{j=1}^{j=n} \frac{P_{jc} - P_{ja}}{d_j n}$$

where n is the number of distance bins, d_j is the upper limit for each bin, P_{jc} is the percentage of correlated pairs with distances between j and $j - 1$, and P_{ja} is the same percentage for all pairs of positions.

Structural model of glucosylceramide binding to the catalytic pocket of GBA

To optimize protein geometry and release local constraints among side-chains, the crystallographic structure of GBA (1OGS) was subjected to three steps of 50 cycles of steepest descent energy minimization using the Deepview program.[21] The molecular structure of glucosylceramide was built using the Corina program (Molecular Networks, GmbH). The lipid chains of the ceramide were

DOI 10.1002/prot

Table II
Amino Acid Changes Identified as Gaucher-Disease-Causing Mutations Mapped in the Substrate-GBA Interaction Region

Residue position	Amino acid replacement (wt → mutant)	Severity	Enzyme activity	Reference[a]
127	Asp → Val	Mild		Ref. 8
178	Pro → Ser	Severe		Ref. 46
179	Trp → Stop	Null		Ref. 47
182	Pro → Leu	Severe	Inactive	Refs. 26 and 7
	Pro → Thr	Unknown		Ref. 2
235	Glu → Gly	Not natural	Nearly inactive	Ref. 47
237	Ser → Pro	Severe		Ref. 29
245	Pro → His	Unknown		Ref. 8
285	Arg → Cys	Unknown		Ref. 2
	Arg → His	Severe		Ref. 48
312	Tyr → Cys	Mild		Ref. 49
313	Tyr → His	Unknown		Ref. 50
341	Ala → Thr	Severe		Ref. 34
342	Cys → Gly	Severe		Ref. 31
	Cys → Arg	Mild		Ref. 51
	Cys → Tyr	Severe		Ref. 8
380	Asp → Asn	Unknown		Ref. 2
	Asp → Ala	Severe		Ref. 52
	Asp → His	Unknown		Ref. 53
382	Asn → Lys	Severe	Reduced to 22%	Ref. 54
391	Pro → Leu	Unknown	Nearly inactive	Refs. 55 and 7
392	Asn → Ile	Severe	Nearly inactive	Ref. 55 and 7
394	Val → Leu	Severe	8.5 times lower	Ref. 56 and 57
396	Asn → Thr	Mild	7 times lower	Ref. 58
397	Phe → Ser	Mild		Ref. 8
398	Val → Leu	Severe		Ref. 59
	Val → Phe	Severe		Ref. 60

In bold: those positions identified as being directly located in the substrate interacting pocket of GBA and whose mutation has been reported as causing Gaucher disease. Mutations in residues located in positions ±1 (related to those directly involved) have been included in this study, as well as other close residues (±3).
[a]When two references are provided for a mutation, the first describes its report as a disease-causing mutation, and the second the expression and enzyme activity studies of the recombinant protein.

assumed not to interact directly with the catalytic pocket, and were replaced by hydrogen atoms to reduce the complexity of calculations. The glucosylceramide structure geometry was optimized using the methods implemented in the MOPAC program.[22] Grid calculations were performed with the Autogrid3 program from the Autodock3 suite,[23,24] generating cubic 75 points and 0.375 Å spacing grid maps centered on the Cβ atom of the active site residue Asp340. Two runs of Autodock3 using the LGA algorithm rendered 200 conformations, which were clustered with an rmsd (root mean square deviation) cut-off of 1 Å for all atoms of each docked solution. Those orientations with the amide-H residue docked within the catalytic pocket were considered unrealistic and discarded. Thus, we selected the optimal docked conformation belonging to the lowest energy of the most populated cluster. Figure plots of protein and ligand structures were generated using PyMOL program (DeLano Scientific, San Carlos CA).

RESULTS AND DISCUSSION

SapC docking model

The aim of our work was to obtain a structural model for the interaction of acid-β-glucosidase (GBA) with both the substrate glucosylceramide (GlcCer) and the co-factor saposin C (SapC) that could explain the effect of mutations on the human GBA sequence responsible for Gaucher disease. Although some of the naturally occurring mutations lie within or close to the active site, and are thus clearly disease related, many others structurally map far from the catalytic pocket and their effects remain unexplained. Because of the fact that saposin C binding to GBA is necessary for enzyme activation, we used the solved structures of both polypeptides to build a protein–protein complex capable of structurally explaining the observed loss of enzyme activity. Figure 1(A) (right) shows a cluster of residues (green) concentrated on one side of the SapC closed structure surface[16] with high values (more than 0.7) compared to GBA in terms of the correlated mutation index.[20] Although the calculation of the correlated mutation index in this particular case is in the limits of the statistical significance,

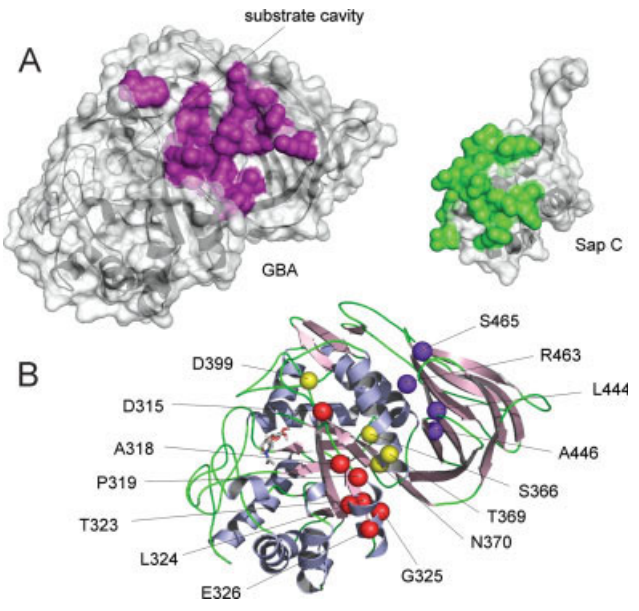

Figure 1
*(**A**) Location of the correlated positions on the SapC (right) and GBA (left) surfaces. Residues in the Saposin C (closed) structure with higher correlation values (>0.7) related to GBA are shown in green (Positions 9, 20, 22, 25, 56, 57, 59, 60, 62–67, 69, 70, and 74 of the human Sap C sequence). Residues in GBA with correlation values >0.7 related to the above-indicated Sap C residues are shown in purple (Positions 314, 317, 318, 348, 358, 362, 365, 366, 369, 370, 372, 373, 441, 443–445, 463, 464, and 487). (**B**) Position of residues whose mutation leads to Gaucher disease; located in the same area as the GBA correlated ones. Residues are colored according to their grouping in Region I (red; mainly TIM barrel helix 6 and the preceding loop), II (yellow; TIM barrel helix 7 and residues in proximity), and III (purple; Ig-like domain).*

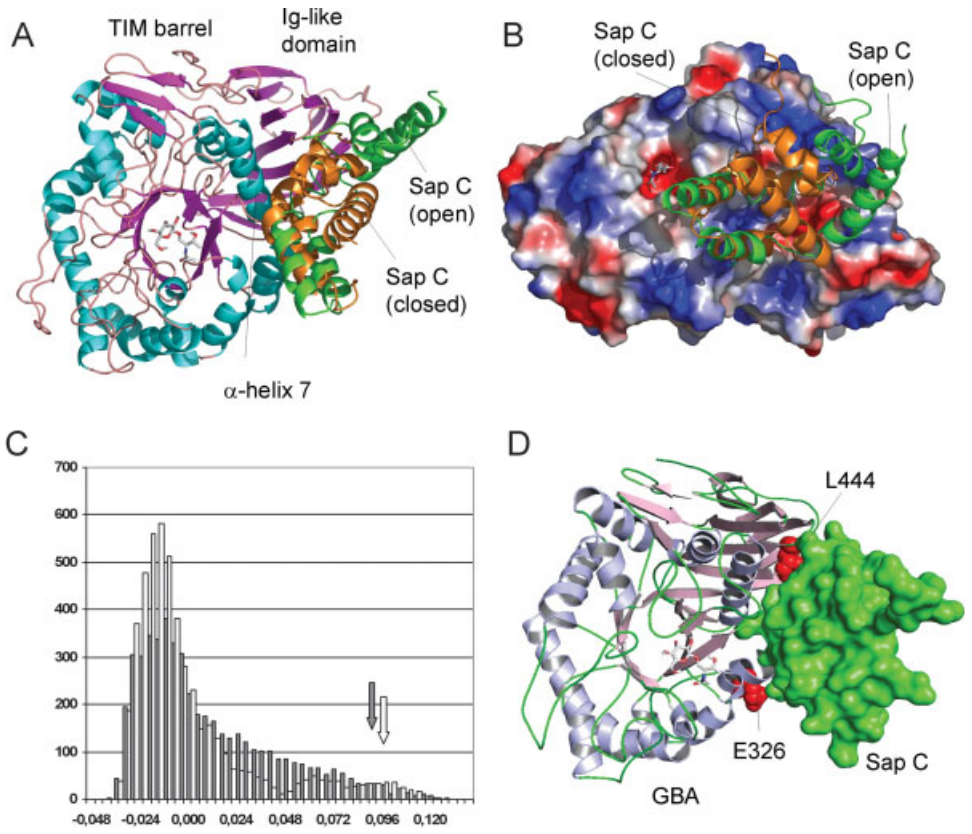

Figure 2

*Docking model for the SapC/GBA interaction. (**A**) Plot of the SapC open (green) and closed (gold) structures located in the GBA region defined by the TIM barrel helix 7 and surrounding structures (TIM barrel helix 6 and Ig-like domain). The position of the GlcCer substrate has been included for clarity. (**B**) Interaction site for SapC open (green) and closed (gold) structures on the surface of the GBA molecule, showing close proximity of SapC to the GBA active site, indicated by the presence of the docked GlcCer substrate. (**C**) Validation of the interaction model. Distribution of the harmonic average factor values (X_d) obtained for 5×10^3 alternative docking solutions. The y-axis represents the number of docking solutions corresponding to each X_d value. Black and white arrows indicate the X_d values calculated for the proposed docking models of open and closed Sap C conformations, respectively. (**D**) Position of the Gaucher disease-causing mutations Leu444Pro and Glu326Lys (red spheres) on both sides of the GBA interaction site and in close proximity to the SapC molecule (green surface), according to the proposed docking model.*

since pairs of sequences from only eight different species were used, the concentration of high-value residues in the same side of SapC structure indicates that the results are consistent with the presence of delimited interaction patches. The left panel of Figure 1(A) shows those residues in GBA (violet) with a higher correlation index related to residues in the above indicated SapC surface cluster. As with SapC, all of the residues in GBA are structurally concentrated in a local cluster, situated around helix 7 of the TIM barrel domain and in proximity to the TIM barrel/Ig-like domains interface. Both patches of residues in the GBA and SapC surfaces are proposed to play a role in protein–protein interactions according to the correlated mutation hypothesis.

Figure 1(B) shows the location of some residues in the GBA structure that are known to be responsible for Gaucher disease when mutated. It is worth noting that they are precisely situated in the putative interaction surface deduced from the correlated mutation analysis. Residues are grouped in three regions. Region I includes residues Asp315, Ala318, Pro319, Thr323, Leu324, Gly325, and Glu326, located in the TIM barrel α helix 6 or in the immediately preceding loop after the β strand 6 (indicated in red). Region II encompasses residues Ser366, Thr369, Asn370, and Asp399, located in the TIM barrel α helix 7 or in its proximity (yellow). Four additional residues, located in the Ig-like GBA domain, Leu444, Ala446, Arg463, and Ser465 conform the Region III (violet).

Figure 2(A,B) shows the model for closed[16] and open[12] structures of SapC in the surface of GBA, obtained using the Hex program[17] for rigid protein–protein docking. The SapC open structure is shown in green while the closed structure is shown in gold. Both SapC conformations bind to the same GBA region, in the vicinity of TIM barrel α helix 7, by the opposite side than that involved in the structural rearrangement leading to the alternative Sap C closed/open structures. Studies on the long-distance influence of the SapC:GBA interaction

DOI 10.1002/prot

on substrate binding and catalysis will be approached using molecular dynamics computational simulations. We evaluated the accuracy of the model using correlated-mutations analysis integrating evolutionary-derived information from multiple sequence alignments as well as structural information obtained from three-dimensional models.[15,20] A weighted harmonic average factor (X_d) was used to measure differences in proximity of correlated residues, indicating positive X_d values for which the predicted interacting patches were closer than the average of all residue population. Correct docking models exhibit higher X_d values than incorrect ones, as has been experimentally confirmed.[15,25] Figure 2(C) shows that the X_d values for the proposed interaction between GBA and both SapC structures (arrows) are among the very highest scores of 5×10^3 alternative solutions used as decoys, indicating a good accuracy for our model.

The effect of several Gaucher disease-causing mutations can be explained using this model, which locates them in the proposed GBA–SapC interaction surface. Table I summarizes these mutations, grouped in the three regions defined above [see Fig. 1(B)]. Region I, those residues in the TIM barrel α helix 6 or in the preceding loop, includes mutations Asp315His, Ala318Asp, Pro319Ala, or Leu324Pro, which modifies the electrostatic or geometric characteristics of the zone, thus affecting correct SapC–GBA interaction. In addition, Region I includes the mutations Thr323Ile, Gly325Trp, Gly325Arg, and Glu326Lys. Thr323 contacts Pro319, Ala320, Ala322, and Leu324, all of them contacting the SapC surface. In addition, Thr323 contacts Arg285 in the vicinity of the GlcCer substrate site (see below), suggesting a signal-transmission role for this residue from the SapC site to the GBA active center. The Thr323Ile mutation can alter its polar contacts to the surrounding residues, thereby modifying enzyme activity levels. The interaction between GBA residue Gly325 and the residue Asp30 in the SapC surface would be substantially altered by the Gly325Trp or Gly325Arg mutations. Mutation Glu326Lys can modify its interactions with Lys321 and Arg329, changing the close interaction of Lys321 to the SapC residue Asp30. Replacements in both the 325 and 326 positions (Gly325Trp and Glu326Lys) render enzymes with a reduced *in vitro* activity (13.9% and 42%, respectively, see Table I).

Region II groups together mutations Ser366Gly, Ser366Asn, Thr369Met, Asn370Ser, Asp399Asn, and Asp399Tyr in the TIM barrel α helix 7. As this helix is located in a hinge region between the TIM barrel and the Ig-like domains, thus connecting the SapC binding site to the enzyme active center in the middle of the barrel sheets, these mutations are supposed to modify the signal transmission between both sites. An example is Ser366, which contacts Trp378, Leu314, and Asn370 in a cluster situated between the SapC interacting site (external face of helix 7) and TIM barrel β sheets 6 and 8, both involved in active site conformation. This could provide an additional explanation for the pathogenic mechanism underlying the most prevalent GD mutation, Asn370Ser, for which only hypotheses have been posited thus far.

Mutations in Region III, found on the surface of the GBA Ig-like domain, include Leu444Arg, Leu444Pro, Ala446Pro, Arg463Cys, Arg463Gln, or Ser465del. This third region closes the clamp formed by Regions I and II (hinge), thus completing the SapC interaction site. Ala446 points toward the SapC interaction surface, contacts Ile368 and Val447, at the surface of the TIM barrel and the Ig-like domains, respectively, and is involved in the proper structural arrangement of the surrounding area, which includes the SapC interaction surface. Leu444 is located between two Asp residues (443 and 445) that interact with the Lys26 position on the SapC surface. Mutations Leu444Arg and Leu444Pro will disrupt the correct orientation of Asp443 or Asp445, due to electrostatic attraction and changes in local backbone structure, respectively, thus modifying the interaction between GBA and SapC. This provides additional basis for the severe effect of the prevalent Leu444Pro mutation, as well as for the reduced activity of the recombinant enzyme preparations (2–18.1%, Table I). Figure 2(D) shows the position of Leu444 and Glu326 in the GBA structure, located oppositely in the proposed binding site for the SapC molecule. Interestingly, the double mutant allele [Leu444Pro-Glu326Lys] exhibits not only a more severe phenotype than the individual mutations,[26] but also lower enzymatic activity (3.5–8.5%) as shown in expression studies.[7] As both residues are located in complementary sites of the GBA–SapC interaction surface, the effect of the double mutant can now be easily explained as a cooperative effect of both mutations. Besides, data on *in vitro* activity of the Arg463Cys protein (24.5%, Table I) and more appallingly the lack of the Ser465 residue (5.5%, Table I) highlight the importance of proposed closing clamp for correct SapC interaction.

Refined substrate-binding model

The second aim of this work was to generate a refined model for the interaction between the GlcCer substrate and the GBA active center. Although a diagram was previously published by Divr *et al.*,[9] a more detailed structural model of the surrounding area was needed to better understand the enzyme functionality of the disease-causing mutations. Using the published structure of human acid-β-glucosidase (1OGS)[9] and the 3D coordinates for the GlcCer substrate, we built a model for the GlcCer/GBA interaction. To optimize local residue geometry and contacts among side chains, residues in the active site were subjected to standard energy minimization procedures. The initial location of the substrate in the enzyme-accommodating pocket was determined by displacing voluminous side chains of Tyr244 and Tyr313 to

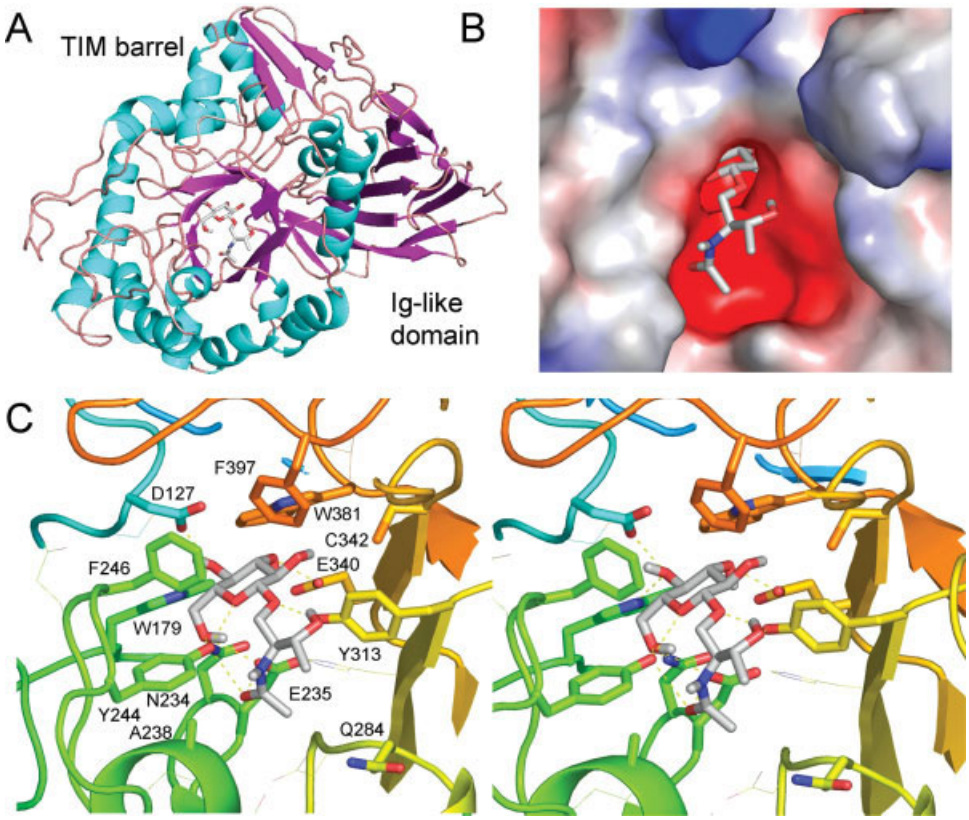

Figure 3

*(**A**) Model of the GlcCer substrate molecule inside the active center of GBA. The position of the TIM barrel and the Ig-like domains are also indicated. (**B**) Inner surface of the substrate active site showing its electrostatic properties. Note the narrow entrance neck of the substrate cavity. (**C**) Stereo-diagram of the substrate/enzyme-docking model. Residues in the active site located close to the substrate are indicated. The ceramide lipid chains of the substrate have been omitted to facilitate the docking computation.*

facilitate placement. Following low-resolution docking steps, these side chains were again located in the upper side of the substrate hole, closing the cavity. Two runs of Autodock3 yielded a series of optimal docking conformations. After eliminating unrealistic conformations (i.e., those with ceramide chains not pointing outside the substrate cavity), we selected the optimal docked conformation belonging to the lowest energy of the most populated cluster. The docking result for GlcCer in the active site of human acid-β-glucosidase is shown in Figure 3(A,B). As expected, the docking solution was located in the same site as that reported for the GBA inhibitor conduritol-β-epoxide.[45]

Figure 3(C) shows in detail the putative contacts of the GluCer molecule with the amino acids surrounding the active center of GBA. Some interesting polar contacts are indicated in this figure. These involve residues that participate in the correct "positioning," or in the catalytic processing, of the Glc substrate. The residues located in the GBA–substrate interaction region, the mutations of which have been associated with different phenotypes of Gaucher disease, are listed in Table II. Asp127 contacts the hydroxyl group of the GlcCer, stabilizing the location of the substrate for catalysis. Thus, its mutation to Val probably alters the proper geometry of the active locus and modifies the enzyme activity. Other acidic residues in the vicinity of the substrate are Glu235, Glu340, and Asp380. The first two have been defined as the catalytic glutamates[9] and the only known mutation in one of them (Glu235Gly) nearly inactivates the enzyme (see Table II).

The Asp380 residue contacts Tyr363 in the GBA α helix 7, which suggests a possible role related to the putative structural signal transmission of the Saposin C binding event to the active center (see below). This might explain the pathogenic effects resulting from the substitution of Asp380 by Asn, Ala, or His. A similar situation might occur with the mutations Arg285Cys or Arg285His. The side chain of Arg285 is directed toward the proposed SapC binding site, interacting additionally with residues Ala318 and Thr323 in the SapC-exposed α helix6. This is also the case for the Ala341Thr substitu-

DOI 10.1002/prot

tion, as the side chain of Ala is directed toward the SapC site and interacts with Trp378. The polar OH group of the Thr amino acid will most likely modify the hydrophobic nature of this structural interaction.

Regarding the active center, the mutation of residues directly contacting OH groups in the substrate, such as Tyr313His, can alter the proper conformation for catalysis, thereby severely decreasing enzyme activity. Other residues in the walls of the substrate cavity located less than 4 Å from the GlcCer molecule include Trp179, Cys342, and Asn382. Although their involvement in substrate catalysis cannot be directly deduced from the docking model, their proximity to the substrate and to the catalytic residues suggests that they play a role in the maintenance of a favorable hydrophobic environment, which would be distorted in the case of the Asn382Lys mutation by the presence of a charged side chain, this leading to an 22%-only active enzyme (Table II).

The mutation of other residues located close to the active center can cause Gaucher disease if the geometry of the region interacting with the substrate is modified. Through its contacts to Trp179, mutation Pro178Ser probably interferes with the correct positioning of the TIM barrel β3 sheet, thereby varying the arrangement of Asn234 and its contact with the substrate. This would then disrupt the hydrophobic core contact between Pro178 and Trp209. Similarly, the change of Pro182 to Leu, in the same hydrophobic core and that completely abolishes the catalytic activity,[7] would alter the local geometry of the enzyme, since this residue contacts the bulky residues Trp184 and Trp209. The change of Ser237 to Pro would render two contiguous Pro residues (Pro236 plus Pro237), thereby causing major local backbone changes rapidly propagated to the substrate site.

Finally, the residues Pro391, Asn392, Val394, Asn396, Phe397, and Val398 are located in a loop connecting the TIM barrel β8 strand to the GBA "Domain I."[9] This loop is located in the upper side of the active site, and all of these residues are in contact, either with the substrate molecule (Phe397) or with residues in the active site. In addition, this loop appears to be involved in membrane interaction due to its proximity to the ceramide-moiety of the substrate. All mutations in these positions (see Table II) would disrupt the normal functioning of this loop. These mutations invariably lead to enzymes with markedly decreased activity, as shown by *in vitro* studies (Table II).

CONCLUSIONS

In the present work, we provide an explanation for the pathogenic effect of mutations at 39 GBA positions. However, more than 90 positions, according to the review by Beutler *et al.*,[8] located outside the SapC:GBA or substrate:GBA interfaces, or the catalytic pocket, remain to be explained.

The evolutionary, structure-based docking model presented in this work not only sheds light on the GBA–SapC–GluCer interactions, but also identifies important structural and/or functional regions in those proteins. This approach could be applied to understand the underlying pathogenicity and the phenotypic outcome of disease-causing mutations beyond the mere analysis of the enzyme catalytic residues.

ACKNOWLEDGMENTS

The authors want to thank Alfonso Valencia not only for his initial encouragement and support on this work, but also for bringing people together. They are also grateful to "biomol-informatics SL (www.bioinfo.es)" for bioinformatics consulting.

REFERENCES

1. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeysinghe S, Krawczak M, Cooper DN. Human gene mutation database: 2003 update. Hum Mutat 2003;21:577–581.
2. Beutler E, Demina A, Gelbart T. Glucocerebrosidase mutations in Gaucher disease. Mol Med 1994;1:82–92.
3. Beutler E, Grabowski GA. Gaucher disease. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. The metabolic and molecular bases of inherited disease, 8th ed, Vol. III. New York: McGraw-Hill; 2001. pp 3635–3668.
4. Theophilus B, Latham T, Grabowski GA, Smith FI. Gaucher disease: molecular heterogeneity and phenotype-genotype correlations. Am J Hum Genet 1989;45:212–225.
5. Chabás A, Cormand B, Grinberg D, Burguera JM, Balcells S, Merino JL, Mate I, Sobrino JA, Gonzalez-Duarte R, Vilageliu L. Unusual expression of Gaucher's disease: cardiovascular calcifications in three sibs homozygous for the D409H mutation. J Med Genet 1995;32:740–742.
6. Abrahamov A, Elstein D, Grosstsur V, Farber B, Glaser Y, Hadashalpern I, Ronen S, Tafakjdi M, Horowitz M, Zimran A. Gaucher's disease variant characterised by progressive calcification of heart valves and unique genotype. Lancet 1995;346:1000–1003.
7. Montfort M, Chabás A, Vilageliu L, Grinberg D. Functional analysis of 13 GBA mutant alleles identified in Gaucher disease patients: pathogenic changes and "modifier" polymorphisms. Hum Mutat 2004;23:567–575.
8. Beutler E, Gelbart T, Scott CR. Hematologically important mutations: Gaucher disease. Blood Cells Mol Dis 2005;35:355–364.
9. Dvir H, Harel M, McCarthy AA, Toker L, Silman I, Futerman AH, Sussman JL. X-ray structure of human acid-beta-glucosidase, the defective enzyme in Gaucher disease. EMBO Rep 2003;4:704–709.
10. Vaccaro AM, Tatti M, Ciaffoni F, Salvioli R, Serafino A, Barca A. Saposin C induces pH-dependent destabilization and fusion of phosphatidylserine-containing vesicles. FEBS Lett 1994;349:181–186.
11. Wilkening G, Linke T, Sandhoff K. Lysosomal degradation on vesicular membrane surfaces. Enhanced glucosylceramide degradation by lysosomal anionic lipids and activators. J Biol Chem 1998;273:30271–30278.
12. Hawkins CA, Alba E, Tjandra N. Solution structure of human saposin C in a detergent environment. J Mol Biol 2005;346:1381–1392.
13. Salvioli R, Tatti M, Scarpa S, Moavero SM, Ciaffoni F, Felicetti F, Kaneski CR, Brady RO, Vaccaro AM. The N370S (Asn370→Ser)

mutation affects the capacity of glucosylceramidase to interact with anionic phospholipid-containing membranes and saposin C. Biochem J 2005;390:95–103.
14. Janin J. Assessing predictions of protein–protein interaction: the CAPRI experiment. Protein Sci 2005;14:278–283.
15. Tress M, de Juan D, Grana O, Gomez MJ, Gomez-Puertas P, Gonzalez JM, Lopez G, Valencia A. Scoring docking models with evolutionary information. Proteins 2005;60:275–280.
16. de Alba E, Weiler S, Tjandra N. Solution structure of human saposin C: pH-dependent interaction with phospholipid vesicles. Biochemistry 2003;42:14729–14740.
17. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins 2000;39:178–194.
18. Pazos F, Olmea O, Valencia A. A graphical interface for correlated mutations and other protein structure prediction methods. Comput Appl Biosci 1997;13:319–321.
19. Vakser IA. Protein docking for low-resolution structures. Protein Eng 1995;8:371–377.
20. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein–protein interaction. J Mol Biol 1997;271:511–523.
21. Guex N, Diemand A, Peitsch MC. Protein modelling for all. Trends Biochem Sci 1999;24:364–367.
22. Stewart JJ. MOPAC: a semiempirical molecular orbital program. J Comput Aided Mol Des 1990;4:1–105.
23. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins 1990;8:195–202.
24. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart E, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 1998;19:1639–1662.
25. Carettoni D, Gomez-Puertas P, Yim L, Mingorance J, Massidda O, Vicente M, Valencia A, Domenici E, Anderluzzi D. Phage-display and correlated mutations identify an essential region of subdomain 1C involved in homodimerization of *Escherichia coli* FtsA. Proteins 2003;5:192–206.
26. Chabás A, Gort L, Diaz-Font A, Montfort M, Santamaria R, Cidras M, Grinberg D, Vilageliu L. Perinatal lethal phenotype with generalized ichthyosis in a type 2 Gaucher disease patient with the [L444P;E326K]/P182L genotype: effect of the E326K change in neonatal and classic forms of the disease. Blood Cells Mol Dis 2005; 35:253–258.
27. Walley AJ, Ellis I, Harris A. Three unrelated Gaucher's disease patients with three novel point mutations in the glucocerebrosidase gene (P266R. D315H and A318D). Br J Haematol 1995;91:330–332.
28. He GS, Grace ME, Grabowski GA. Gaucher disease: four rare alleles encoding F213I. P289L, T323I, and R463C in type 1 variants. Hum Mutat 1992;1:423–427.
29. Germain DP, Puech JP, Caillaud C, Kahn A, Poenaru L. Exhaustive screening of the acid β-glucosidase gene, by fluorescence-assisted mismatch analysis using universal primers: mutation profile and genotype/phenotype correlations in Gaucher disease. Am J Hum Genet 1998;63:415–427.
30. Torralba MA, Perez Calvo JI, Pastores GM, Cenarro A, Giraldo P, Pocovi M. Identification and characterization of a novel mutation c. 1090G > T (G325W) and nine common mutant alleles leading to Gaucher disease in Spanish. Blood Cells Mol Dis 2001;27:489–495.
31. Eyal N, Wilder S, Horowitz M. Prevalent and rare mutations among Gaucher patients. Gene 1990;96:277–283.
32. Eyal N, Firon N, Wilder WS, Kolodny EH, Horowitz M. Three unique base pair changes in a family with Gaucher disease. Hum Genet 1991;87:328–332.
33. Ida H, Rennert OM, Kawame H, Maekawa K, Eto Y. Mutation prevalence among 47 unrelated Japanese patients with Gaucher disease: identification of four novel mutations. J Inherit Metab Dis 1997;20:67–73.
34. Demina A, Beutler E. Six new Gaucher disease mutations. Acta Haematol 1998;99:80–82.
35. Beutler E, Gelbart T, Balicki D, Demina A, Adusumalli J, Elsas L, 2nd, Grinzaid KA, Gitzelmann R, Superti-Furga A, Kattamis C, Liou BB. Gaucher disease: four families with previously undescribed mutations. Proc Assoc Am Physicians 1996;108:179–184.
36. Tsuji S, Martin BM, Barranger JA, Stubblefield BK, LaMarca ME, Ginns EI. Genetic heterogeneity in Type I Gaucher disease: multiple genotypes in Ashkenazic and non-Ashkenazic individuals. Proc Natl Acad Sci USA 1988;85:2349–2352.
37. Shamseddine A, Taher A, Fakhani S, Zhang M, Scott CR, Habbal MZ. Novel mutation. L371V, causing multigenerational Gaucher disease in a Lebanese family. Am J Med Genet A 2004;125:257–260.
38. Beutler E, Gelbart T. Two new Gaucher disease mutations. Hum Genet 1994;93:209–210.
39. Koprivica V, Stone DL, Park JK, Callahan M, Frisch A, Cohen IJ, Tayebi N, Sidransky E. Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease. Am J Hum Genet 2000;66:1777–1786.
40. Uchiyama A, Tomatsu S, Kondo N, Suzuki Y, Shimozawa N, Fukuda S, Sukegawa K, Taki N, Inamori H, Orii T. New Gaucher disease mutations in exon 10: a novel L444R mutation produces a new NciI site the same as L444P. Hum Mol Genet 1994;3:1183–1184.
41. Tsuji S, Choudary PV, Martin BM, Stubblefield BK, Mayor JA, Barranger JA, Ginns EI. A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. New Engl J Med 1987; 316:570–575.
42. Filocamo M, Mazzotti R, Stroppiano M, Seri M, Giona F, Parenti G, Regis S, Corsolini F, Zoboli S, Gatti R. Analysis of the glucocerebrosidase gene and mutation profile in 144 Italian Gaucher patients. Hum Mutat 2002;20:234–235.
43. Hong CM, Ohashi T, Yu XJ, Weiler S, Barranger JA. Sequence of two alleles responsible for Gaucher disease. DNA Cell Biol 1990;9: 233–241.
44. Alfonso P, Rodriguez-Rey JC, Ganan A, Perez-Calvo JI, Giralt M, Giraldo P, Pocovi M. Expression and functional characterization of mutated glucocerebrosidase alleles causing Gaucher disease in Spanish patients. Blood Cells Mol Dis 2004;32:218–225.
45. Premkumar L, Sawkar AR, Boldin-Adamsky S, Toker L, Silman I, Kelly JW, Futerman AH, Sussman JL. X-ray structure of human acid-{β}-glucosidase covalently bound to conduritol-B-epoxide: implications for Gaucher disease. J Biol Chem 2005;280:23815–23819.
46. Choy F, Wei C. Identification of a new mutation (P178S) in an African-American patient with type 2 Gaucher disease. Hum Mutat 1995;5:345–347.
47. Grace ME, Desnick RJ, Pastores GM. Identification and expression of acid β glucosidase mutations causing severe type 1 and neurologic type 2 gaucher disease in non Jewish patients. J Clin Invest 1997;99:2530–2537.
48. Stone DL, Tayebi N, Orvisky E, Stubblefield B, Madike V, Sidransky E. Glucocerebrosidase gene mutations in patients with type 2 Gaucher disease. Hum Mutat 2000;15:181–188.
49. Latham TE, Theophilus BDM, Grabowski GA, Smith FI. Heterogeneity of mutations in the acid β-glucosidase gene of Gaucher disease patients. DNA Cell Biol 1991;10:15–21.
50. Cormand B, Vilageliu L, Balcells S, Gonzalez-Duarte R, Chabás A, Grinberg D. Two novel (1098insA and Y313H) and one rare (R359Q) mutations detected in exon 8 of the β-glucocerebrosidase gene in Gaucher's disease patients. Hum Mutat 1996;7:272–274.
51. Choy FYM, Vaags A, Wong K, Macgregor D, Fernandez B, Prasad C. Identification of two novel mutations. L105R and C342R, in type I Gaucher disease Clin Genet 2002;61:229–231.
52. Walley A, Harris A. A novel point mutation (D380A) and a rare deletion (1255del55) in the glucocerebrosidase gene causing Gaucher's disease. Hum Mol Genet 1993;2:1737–1738.

 DOI 10.1002/prot

53. Pomponio RJ, Cabrera-Salazar MA, Echeverri OY, Miller G, Barrera LA. Gaucher disease in Colombia: mutation identification and comparison to other Hispanic populations. Mol Genet Metab 2005; 86:466–472.
54. Tang NL, Zhang W, Grabowski GA, To KF, Choy FY, Ma SL, Shi HP. Novel mutations in type 2 Gaucher disease in Chinese and their functional characterization by heterologous expression. Hum Mutat 2005;26:59–60.
55. Cormand B, Grinberg D, Gort L, Chabás A, Vilageliu L. Molecular analysis and clinical findings in the Spanish Gaucher disease population: putative haplotype of the N370S ancestral chromosome. Hum Mutat 1998;11:295–305.
56. Theophilus BDM, Latham T, Grabowski GA, Smith FI. Comparison of RNase A, a chemical cleavage and GC-clamped denaturing gradient gel electrophoresis for the detection of mutations in exon 9 of the human acid β-glucosidase gene. Nucl Acid Res 1989;17:7707–7722.
57. Verghese J, Goldberg RF, Desnick RJ, Grace ME, Goldman JE, Lee SC, Dickson DW, Rapin I. Myoclonus from selective dentate nucleus degeneration in type 3 Gaucher disease. Arch Neurol 2000;57:389–395.
58. Amaral O, Pinto E, Fortuna M, Lacerda L, Miranda M. Type 1 Gaucher disease: identification of N396T and prevalence of glucocerebrosidase mutations in the Portuguese. Hum Mutat 1996;8:280–281.
59. Seeman P, Finckh U, Hoppner J, Lakner V, Liebisch I, Grau G, Rolfs A. Two new missense mutations in a non-Jewish Caucasian family with type 3 Gaucher disease. Neurology 1996;46:1102–1107.
60. Stone DL, van DO, de KJ, Gaillard J, Niermeijer MF, Willemsen R, Tayebi N, Sidransky E. Is the perinatal lethal form of Gaucher disease more common than classic type 2 Gaucher disease? Eur J Hum Genet 1999;7:505–509.