Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay–Sachs disease

Ivo Tews, Anastassis Perrakis, Amos Oppenheim[1], Zbigniew Dauter, Keith S. Wilson and Constantin E. Vorgias[2]

Chitin, the second most abundant polysaccharide on earth, is degraded by chitinases and chitobiases. The structure of *Serratia marcescens* chitobiase has been refined at 1.9 Å resolution. The mature protein is folded into four domains and its active site is situated at the C-terminal end of the central $(\beta\alpha)_8$-barrel. Based on the structure of the complex with the substrate disaccharide chitobiose, we propose an acid-base reaction mechanism, in which only one protein carboxylate acts as catalytic acid, while the nucleophile is the polar acetamido group of the sugar in a substrate-assisted reaction. The structural data lead to the hypothesis that the reaction proceeds with retention of anomeric configuration. The structure allows us to model the catalytic domain of the homologous hexosaminidases to give a structural rationale to pathogenic mutations that underlie Tay–Sachs and Sandhoff disease.

European Molecular Biology Laboratory Hamburg Outstation Notkestraße 85, 22603 Hamburg, Germany
[1]Department of Molecular Genetics The Hebrew University-Hadassah Medical School Jerusalem 90101 Israel.
[2]From September 1996 C.E.V will move to: Athens University, Biology Department, Biochemistry Laboratory, Panepistimiopoli, Kouponia, 15701 Athens, Greece

Correspondence should be addressed to C.E.V.

Chitin is composed of *N*-acetylglucosamine units (β-1,4 linked 2-acetamido-2-deoxy-glucopyranosyl, NAG). After cellulose, chitin is the most abundant carbohydrate polymer. Chitin plays a significant morphogenetic and structural role in fungi, insects, crustaceans and plants and is also an important nutrient for bacteria.

Chitin metabolizing enzymes are glycosyl hydrolases, which split glycosidic linkages between adjacent sugar residues. Glycosyl hydrolases are classified into 45 families on the basis of amino acid sequence[1]. To date, three-dimensional structures are known for representatives of 22 families[2]. Within the families, variations in chain length and hence in domain structure are observed. Many have an eight-stranded $(\beta\alpha)_8$-barrel as the catalytic domain. Structural differences reflect variation in substrate specificity, anomeric specificity (such as α-1,4 or β-1,4 linkages) and retention or inversion of the anomeric configuration. All members of any one family display the same mechanistic stereochemistry.

The chitinolytic glycosyl hydrolases are found in all organisms in which chitin is a structural component. Plants, lacking an immune system, use chitinases as a defence mechanism[3]. Bacteria use chitin as a food source and their chitinolytic systems are involved in the natural recycling of chitin[4]: these enzymes are of biotechnological relevance especially as antifungal agents. Three-dimensional structures are known for prokaryotic and eukaryotic chitinases falling into families 18 (refs 5–8) and 19 (ref. 9). Bacterial chitobiases (EC 3.2.1.30) belong to family 20; there are no previous crystal structures available for representatives of this family. *Serratia marcescens* is an ideal organism for structural and mechanistic studies of these enzymes: it is one of the most active chitinolytic bacteria[10], producing enzymes with high activity over a wide pH range[11]. *S. marcescens* possesses a set of chitinases that hydrolyse chitin to the disaccharide chitobiose (diNAG) which is then hydrolysed to NAG by chitobiase. The degradation products are then taken up by the cell.

Family 20 also contains other enzymes, the hexosaminidases, which hydrolyse G_{M2}-type gangliosides in higher organisms. Deficiencies in these enzymes lead to accumulation of gangliosides[12], resulting in gangliosidoses such as Tay–Sachs or Sandhoff disease. Numerous mutations in human hexosaminidase genes have been identified and classified according to the severity of their phenotype[13]. Structural information presented here on a bacterial member of family 20, chitobiase, allows modelling of the catalytic domain of hexosaminidase α, providing a first step towards an understanding of these diseases at the structural level.

Structure of chitobiase

Chitobiase was constitutively expressed in *Escherichia coli*. The gene sequence predicts a 98,500 M_r protein[14]. The newly synthesised protein is transported to the periplasm and undergoes proteolytic cleavage of its leader peptide. The crystal structure of the protein was determined by multiple isomorphous replacement, and refinement statistics reflect good data quality (Table 1).

Chitobiase is a monomeric protein with four domains (Fig. 1). The overall dimensions are ~90×80×60 Å, and three disulphide bridges stabilize the structure. There are no cofactors, metals or other ligands in the native enzyme. On one side, opposite to that containing the active site, the enzyme is essentially flat (in the plane of the view shown in Fig. 1): this could represent a potential membrane binding surface. Domains I, II and III form a compact body from which

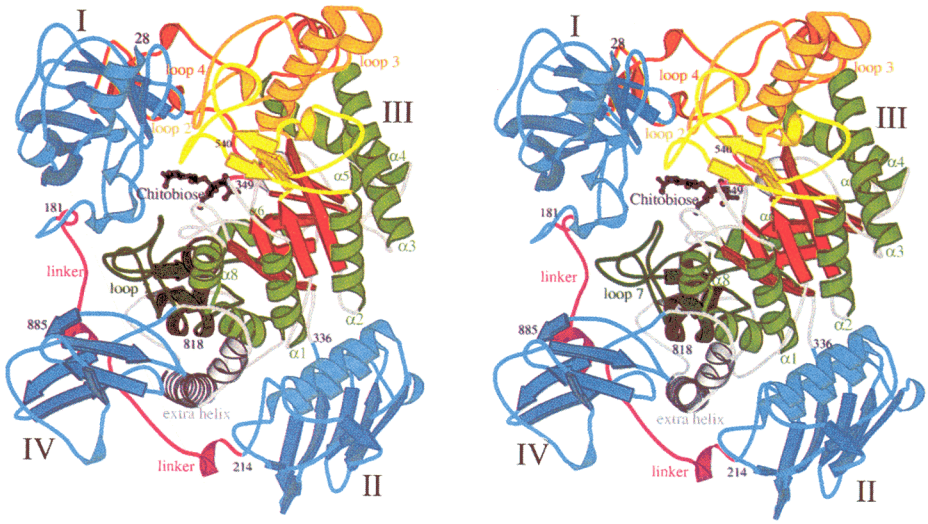


Fig. 1 Cartoon diagram[76] of chitobiase complexed with the natural substrate diNAG. The substrate binding pocket is open at the back in the orientation shown. Roman numerals indicate domains, Arabic numerals indicate amino acid positions. Two key active site residues are shown (Arg 349, Glu 540). Domain I starts with a 2½ turn helix pointing into solvent. A long linker, magenta, connects domains I and II. The Cαs of Met 223 at the beginning and Asp 335 at the end of Domain II are separated by only 4.8 Å. Domain III, the eight stranded $(\beta\alpha)_8$-barrel, is shown side-on. The eight β-strands, in red, are surrounded by seven α-helices, green. Helix α8 is kinked after two turns. Loops inserted in the barrel are coloured yellow (loop 2), orange (loop 3) and ochre (loop 4): they are in close contact with domain I. Loop 7 (dark green) replaces the helix α7 that would be found in a normal $(\beta\alpha)_8$-barrel . The final features of this domain (grey) are a β-strand and a kinked seven turn extra helix winding around the barrel. Domain IV, an immunoglobulin-like fold is in contact with a helix in the linker region.

the C-terminal domain IV protrudes. Connections between domains are short except for an elongated linker containing two helices which connects domains I and II. All four domains have similar average temperature factors and are well ordered.

The N-terminal domain I (residues 28–181) comprises two β-sheets. It has a highly ordered hydrophobic core leading to a compact structure. A two-and-a-half turn helix precedes the first β-strand; the N terminus of the helix points into the solvent and is an obvious cleavage site for the 27 amino-acid long leader peptide. Domain I contains a long loop, stabilized by a disulphide following the first β-strand, which interacts with domain III. Domain I is very similar to the cellulose binding domain (CBD) of the cellulase from *Cellulomonas fimi*[15], with identical overall topology and similar β-strand length. Two differences are the presence of the two-and-a-half turn helix and of the long loop between β-strands 1 and 2 in chitobiase. In cellulases, the CBD is often linked to the core domain(s) by a flexible peptide, but in chitobiase there are well ordered hydrophobic interactions between domains I and III.

Domain II (residues 214–335) has α+β topology. Two regular α-helices, buried inside the protein, are tilted about 30° with respect to an exposed, seven-stranded sheet. This domain shows topological similarities to other known proteins (such as the *Serratia* metalloproteinase, serralysin[16]). The exposed sheet and the parallel arrangement of the buried helices appear to be a novel structural feature.

All the other domains are organized around domain III (residues 336–818), which is a $(\beta\alpha)_8$-barrel. The active site is on the C-terminal end of the barrel. There are significant deviations from the basic $(\beta\alpha)_8$-barrel motif. First, there are only seven α-helices. Helix α7 is absent and is replaced by the 67 amino-acid long loop 7. Helix α5 is reduced to one turn. Second, there are long loops at the C-terminal end of the barrel, which occur after β-strands 2, 3 and 4. The loops are between 51 and 58 amino acids in length, contain β-sheets and helical segments, and

Table 1 Crystallographic analysis

	Native	Gold	Platinum	Chitobiase
Diffraction limit (Å)	15–1.9	15–2.0	15–2.4	15–2.0
Unique reflections	83,596	75,629	39,502	66,462
Overall R_{merge} %[1]	6.2	6.2	6.8	8.4
Overall completeness %	99.9	96.2	99.9	99.7
Overall I/σ	15.8	12.0	14.2	14.8
Overall redundancy	5.5	3.7	3.9	4.1
HR[2] shell (Å)	1.88–1.85	2.01–1.96	2.45–2.37	2.06–2.02
HR R_{merge} %[1]	27.6	17.7	12.0	23.7
HR completeness %	99.8	95.1	100	99.9
HR I/σ	3.2	4.1	8.4	3.4
R_{Deriv} %[1]	-	10.1	17.7	10.1
PhP[3] (iso / ano)	-	2.4 / 2.1	1.5 / 1.4	-
R.m.s. (bonds 1–2/1–3)[4]	0.022/0.032	-	-	0.019 /0.029
R_{factor}[5]	14.6	-	-	15.2

[1] $R_{merge}=\Sigma|I-I_{mean}|\Sigma|I|$; $R_{deriv}=\Sigma|\boldsymbol{F}_P-\boldsymbol{F}_{PH}|/\Sigma|\boldsymbol{F}_P|$, where $\boldsymbol{F}_P$=protein structure factor amplitude, $\boldsymbol{F}_{PH}$=derivative structure factor amplitude.
[2] HR=high resolution.
[3] Phasing power (PhP)=r.m.s. ($\boldsymbol{F}_H/E$), where $\boldsymbol{F}_H$=heavy atom structure factor, E = residual lack of closure
[4] R.m.s.=root mean square deviations from ideal values.
[5] $R_{factor}=\Sigma|\boldsymbol{F}_{obs}-\boldsymbol{F}_{calc}|/\Sigma\boldsymbol{F}_{obs}$.

interact with domain I. Third, an additional β-strand and a long, seven-turn helix complete domain III. This helix, kinked after four turns, winds around the barrel and interacts with domain II. This feature of domain III leads directly into domain IV. The sequence of this domain is roughly twice as long as that of the classic $(\beta\alpha)_8$-barrel of triose phosphate isomerase[17].

Domain IV, like domain I, comprises two β-sheets but is only 67 amino acids in length (residues 819–885). The C-terminal valine is in the β-conformation and is buried inside the protein. Domain IV is similar to the immunoglobulin β-sandwich. The domain superimposes well on the structure of human growth hormone receptor[18], but there is a small helical segment in chitobiase instead of one of the seven strands that occurs in the receptor.

The complex and active site architecture

The structure of a complex of chitobiase with a substrate may be expected to shed light on various aspects of the enzyme's catalytic mechanism, including reaction kinetics[19] and whether the reaction proceeds with retention or inversion of the anomeric configuration of substrate[20–22]. To obtain a complex, a crystal of chitobiase was exposed for less than ten minutes to crystallization buffer containing diNAG (chitobiose), the enzyme's natural substrate. In the crystal complex, the

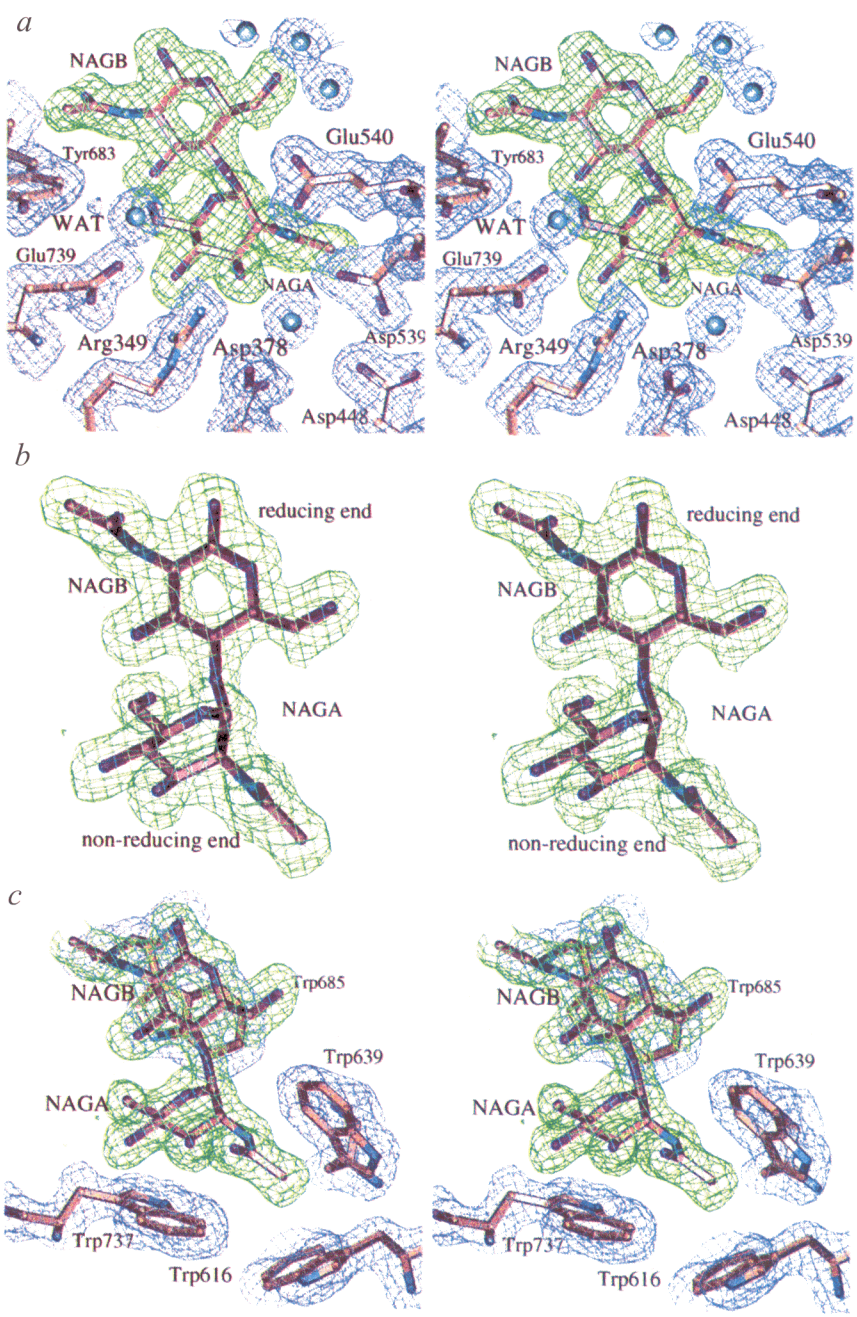
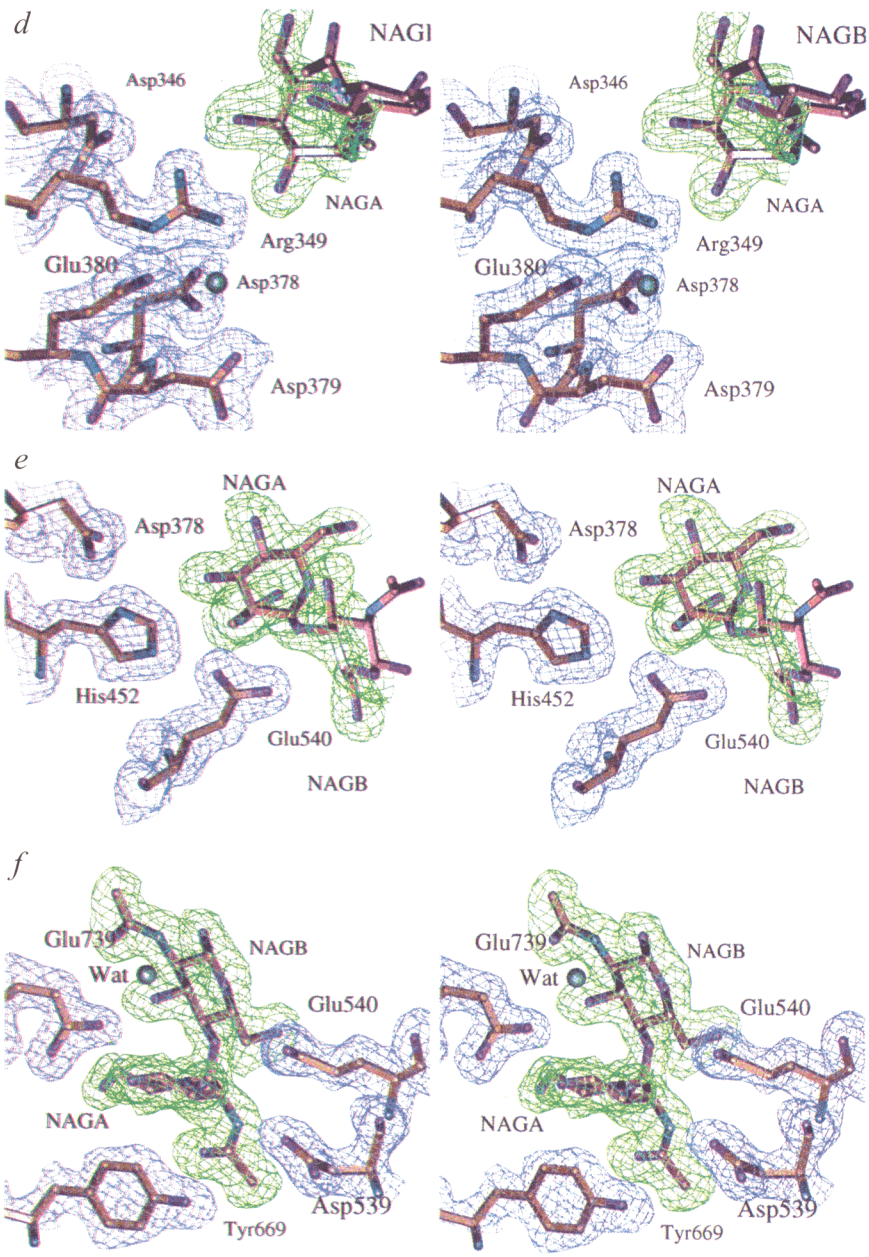


Fig. 2 Substrate binding of chitobiase. *a*, In the boot-shaped active-site pocket, the inner sugar NAGA, bound to Arg 349, is surrounded by protein groups and the outer NAGB is more exposed to solvent. Note the distortion of NAGA and the position of the catalytic Glu 540 binding to the glycosidic oxygen, which is involved in the bond to be cleaved. The water we believe to be involved in hydrolysis is marked WAT. *b*, diNAG is distorted. NAGB is in the $^4C_1$ chair; NAGA is in a conformation close to the energetically unfavourable 4-sofa. The twist around the glycosidic linkage disrupts the hydrogen bond between O5A and OH3B. *c*, NAGB stacks against Trp 685, NAGA against Trp 737. The essentially planar *N*-acetyl group of NAGA is stabilized by Trp 616 and Trp 639. A very small movement of these aromatic residues relative to the native structure creates space for the incoming sugar. *d*, *next page*, the substrate is tightly anchored through hydrogen bonds from the OH3A and OH4A groups of the non-reducing sugar NAGA to Arg 349. Arg 349 is held in place by polar interactions with Asp 346, Asp 378, Asp 379, Glu 380 and a water (blue sphere), where the terminal amino and imino groups hydrogen bond to Asp 346 (Nη1), Glu 380 (Nε), and a water (Nη2) coordinated by Asp 379 and Glu 380. *e*, *next page*, Glu 540 binds to the glycosidic linkage, and is a candidate for the catalytic acid. Oε1 accepts a hydrogen bond (2.7 Å) from His 452, which is itself hydrogen bonded to Asp 378. Oε2 is at a distance of 2.9 Å from the glycosidic O4B, an ideal length for a hydrogen bond and suggesting it is protonated. *f*, *next page*, the *N*-acetyl-group of the non-reducing sugar ring may act as nucleophile stabilizing the reaction intermediate, probably through formation of a covalent oxazoline intermediate. It is held in place by hydrophobic interactions (Fig. 2*c*) and hydrogen bonds from acetamido-*N*2A to Asp 539 and from acetamido-O7A to Tyr 669. As a result, O7A is very close (3.0 Å) to the anomeric centre C1A. A water hydrogen bonded between OH5B and Glu 739 is a candidate for the nucleophile in the hydrolysis of the glycosidic bond. Stereograms of the $2\boldsymbol{F}_{obs}-\boldsymbol{F}_{calc}$ electron density map were calculated with phases from the final refined model contoured at 1σ.

substrate is not cleaved and is present at almost full occupancy as judged from the crystallographic analysis. This is the first complex of a glycosyl hydrolase with its natural substrate. The stability of this apparently abortive complex may be due to the high salt content combined with restricted flexibility in the crystal lattice, presenting unfavourable conditions for catalysis. Soak times of a few hours reduced the resolution limit to less than 5 Å and after one day there was essentially no observable diffraction. This may suggest that the present structure is a 'pre-Michaelis' complex with induced conformational changes unable to occur due to the effects of the crystal packing and the short soak times. Induced changes are presumed to occur slowly and lead to disorder of the crystal. Attempts at co-crystallization were unsuccessful.

In the complex, there is an intact glycosidic linkage between sugar moieties in the –1/+1 sites on either side of the cleavage point. The two sugar residues are referred to in this report as the non-reducing NAGA in the –1 site, and the reducing NAGB in the +1 site. The C and O atoms of each sugar ring carry the suffix A or B as appropriate. The active site lies at the centre of the convex side of the enzyme, at the C-terminal end of the $(\beta\alpha)_8$-barrel. diNAG binds in a boot-shaped pocket (Fig. 2). The two sugar planes are tilted around their glycosidic linkage by ~90° with respect to one another (Fig. 2*a*, *b*). The intramolecular hydrogen bond between O5A and OH3B that stabilizes the extended chain conformation in β-1,4 linked glycosides is thus disrupted. Some other major features of how the substrate pocket binds to diNAG are worth noting. First, diNAG is distorted when bound to the enzyme: the outer sugar NAGB is in the typical $^4C_1$ chair, while the inner sugar, NAGA, is in a conformation close to the energetically unfavourable 4-sofa (Fig. 2*b*). Second, four tryptophans line part of the binding pocket complementary to the hydrophobic surfaces of the sugar rings (Fig. 2*c*). Third, the non-reducing NAGA is 'docked' to the enzyme by hydrogen bonds from OH3A and OH4A to Arg 349, itself stabilized by polar interactions to other protein groups. Arg 349 sits at the base of the binding pocket (Fig. 2*d*). Fourth, Glu 540 forms a hydrogen bond to the glycosidic oxygen (Fig. 2*e*). Fifth, the acetamido group of the non-reducing sugar ring NAGA is bent towards the anomeric C1A. The distortion is stabilized by hydrophobic interactions (Fig. 2*c*), and by polar interactions with Asp 539 and Tyr 669 (Fig. 2*f*).



An overview of the diNAG complex (Fig. 3)—combining all the features in Fig. 2—shows the hydrogen bonds between substrate and protein. All polar atoms of the non-reducing sugar ring NAGA—except O5A—make direct hydrogen-bonds with the protein. The reducing sugar ring NAGB, accessible from the surface, makes only indirect hydrogen bonds with protein groups through bridging water molecules and has higher temperature factors than NAGA. In addition the other important interactions are evident, namely the packing of tryptophans around the hydrophobic surfaces of the sugar, the vital role of Arg 349 in anchoring NAGA, the position of Glu 540 and of the N-acetyl moiety of NAGA.

The carboxylate residue acting as the catalytic acid in chitobiase is Glu 540, which binds (distance of 2.9 Å) to the glycosidic linkage (Fig. 2*e*). As may be expected, Glu 540 is invariant among family 20 enzymes. In chitobiase there is

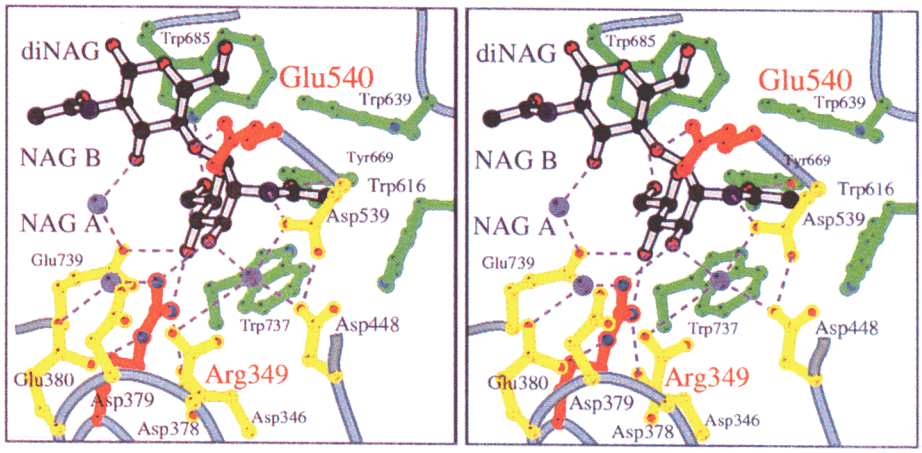


Fig. 3 Stereo view of the hydrogen bond scheme in the substrate binding. There are 16 residues and several water molecules involved. The residues lie at the C-terminal end of the $(\beta\alpha)_8$-barrel and each of the eight loops connecting strands and helices contributes at least one residue. Not all of these residues are shown. Aromatic residues are in green, polar residues are yellow. Glu 540 and Arg 549, playing a key role in catalysis and binding are highlighted in red.

no second protein carboxyl group at an appropriate distance for either a retaining or for an inverting mechanism. However, due to the distortion of NAGA, its acetamido O7A atom lies 3.0 Å from C1A and is ideally poised to act as the nucleophilic base.

Some of the general structural features underlying hydrolysis by chitobiase are comparable to those seen in chitinases. Chitinases, however, have a groove to which the polysaccharide chain can bind randomly[2,5–9], while the active site in chitobiase is a pocket into which the substrate docks. This can be related directly to the endolytic action of the chitinases which need to bind an extended polymer or oligomer along the active site. In contrast chitobiase removes a single monosaccharide from the non-reducing end of a small oligomer.

Proposed mechanism

Based on the crystal structure of the diNAG complex and previous biochemical studies[21,22] we propose that chitobiase uses an acid-base reaction mechanism with Glu 540 as the catalytic acid and with the N-acetyl group of the substrate replacing the enzymatic nucleophilic base found in other glycosyl hydrolases. From the structures, it is evident that there are minimal changes in the conformation of the protein on forming the diNAG complex; only a small movement of aromatic residues is required for them to pack closely against hydrophobic surfaces of the sugar. The two sugar residues in the complex are at a totally different angle to one another compared to the conformation observed in the small-molecule structure of diNAG[23] and that observed for β-1,4 linkages in cellulose or chitin.

The more tightly bound non-reducing sugar NAGA is distorted towards the expected transition state from the usual $^4C_1$ chair conformation into a sofa-like structure, and its acetyl group is bent below the ring plane so that the acetamido carboxyl comes close to the anomeric C1A. While the resolution (2.0 Å) of the present analysis shows clearly that NAGA is not in the $^4C_1$ chair conformation, it is not sufficient to distinguish the 4-sofa from other closely related conformations, or indeed from a mixture of these (Fig. 4). The O1B oxygen at the reducing end of diNAG is a mixture of α and β anomers, as expected. In effect, the conformation of the disaccharide is such that NAGB, which would in chitin be described as being β-1,4 equatorial to NAGA, is more axial. This conformation is energetically more favourable for a leaving group. The distortion also makes the anomeric C1A planar.

The oxygen in the glycosidic linkage is hydrogen bonded to the protonated carboxyl group of Glu 540, the acid in the acid catalysis (Fig. 2*e*). As the first step in the reaction, we envision that this glutamate transfers its proton to the glycosidic oxygen. The C1A–O4B bond between the two sugars would then be split. The departure of NAGB would leave a potential carbonium

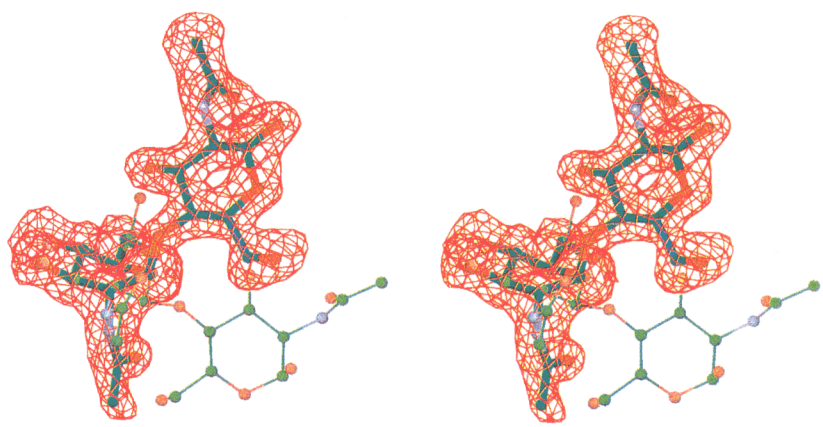


Fig. 4 Sugar distortion. The final, refined model of diNAG, shown in thick bonds, is overlayed on a 5σ unbiased electron density difference map calculated using the differences between diNAG in the complex and native chitobiase data, with phases calculated from the native chitobiase model. To show the distortion, the non-reducing ring NAGA from the small molecule model of diNAG, shown in thin bonds, is superimposed on that of the present model. The non-reducing rings superimpose well, but their conformations are substantially different. The half-chair conformation of NAGA causes the reducing NAGB ring to be in a different position. The glycosidic linkage is more twisted in the chitobiase-diNAG complex.

ion on NAGA. In practice this intermediate can be stabilized by the *N*-acetyl group of NAGA itself, either through formation of a covalent bond between O7A and C1A resulting in an oxazoline ring, or by electrostatic stabilization of the oxocarbonium ion by the partial negative charge of the carbonyl oxygen of the acetamido group. In both cases, this would stabilize the reaction intermediate for a sufficient time for NAGB to diffuse out. The departure of NAGB would allow its replacement by an incoming water molecule on the same side of the ring, which attacks C1A to complete the hydrolysis. This proposed double displacement mechanism results in retention of anomeric configuration. The best candidate for the incoming water molecule is that located between Glu 739 and the reducing sugar ring in the substrate complex (Fig. 2*a,f*). Glu 540 could activate this water by abstracting a proton from it.

Thus from the structural data it appears that the reaction probably proceeds with retention of configuration. This is supported by preliminary kinetic data which provide direct evidence for a retaining mechanism (S. Armand, C.E. V. and B. Henrissat, unpublished results). Furthermore, the homologous hexosaminidases, classified in the same glycosyl hydrolase family, degrade substrate with an overall retention of configuration[24]. As stated above, enzymes from the same glycosyl hydrolase family in all known cases catalyse the reaction with the same mechanism: this can reasonably be assumed to be true for chitobiase and hexosaminidase from family 20.

Enzymatic hydrolysis of a saccharide substrate with overall retention of the configuration implies the participation of a nucleophilic/negatively charged residue resulting in the formation of a glycosyl enzyme intermediate which is subsequently hydrolysed by a water molecule. In contrast to many retaining glycosyl hydrolases, there is no such group on chitobiase: the intermediate stabilization would then have to be provided by the O7 atom of the acetamido group of the substrate itself.

Retention in chitobiase through anchimeric assistance is consistent with similar evidence for other retaining glycosyl hydrolases or transglycosylases through the *N*-acetyl group at the C2 position or other substituents. Examples include hevamine[8], soluble lytic transglycosylse[25] (muraminidase) and probably also goose lysozyme[26]. For hevamine this was based on the structure of an enzyme–inhibitor complex which mimics the reaction intermediate rather than a natural chitin-like oligomer[27]. Hen egg white lysozyme would appear to be an exception since it is believed to act through an electrostatic stabilization of the intermediate oxocarbonium ion primarily by Asp 52. However, even in the case of a hen egg white lysozyme, the participation of the C2 acetamido group cannot be totally excluded[28].

Chitobiases and hexosaminidases

The structural information above allows productive comparison with other family 20 members, notably the hexosaminidases. Chitobiases hydrolyse β-1,4 linked NAG moieties, whereas hexosaminidase hydrolyse the β-1,4 linkage between *N*-acetyl-galactosamine and galactosamine. The substrates for mammalian hexosaminidases are G_{M2} gangliosides (*N*-acetylgalactosaminyl-β-1,4- (*N*-acetylneuraminyl-α-2,3)-galactosyl-β-1,4-glycosyl-β-1,1-ceramide). Sequence identities between *Serratia* chitobiase and other glycosyl hydrolase family 20 enzymes, either bacterial chitobiases or eukaryotic hexosaminidases, range between 16–54%. The sequence data imply that the relationships between family 20 members form three distinct branches in a phylogenetic tree, plus a fourth intermediate group. The first group contains three bacterial

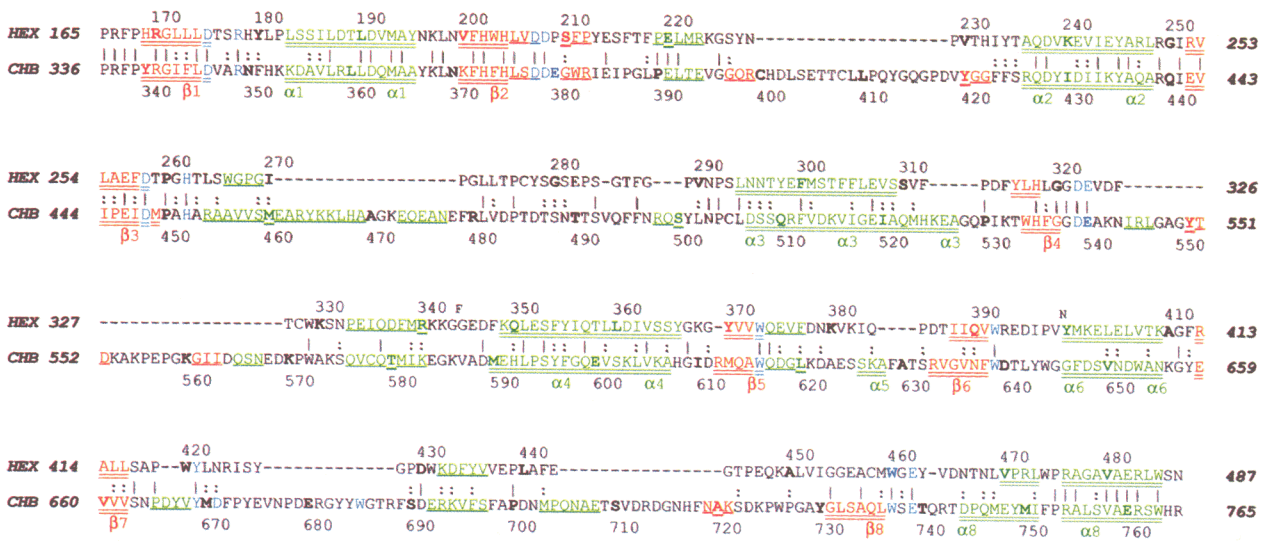


Fig. 5 Sequence alignment for the catalytic domain III of chitobiase and human hexosaminidase α-chain. In contrast to a conventional approach, the positions of insertions and deletions were adjusted to align secondary structural elements predicted from multiple sequence alignments for hexosaminidases with their known positions in chitobiase (see text). A similar alignment is obtained using the β-chain (not shown). Underlining indicates α-helices (green) and β-strands (red). Double underlining indicates central elements of the $(\beta\alpha)_8$-barrel fold. Residues coloured blue are important for active site formation (see Fig. 3). Identities are marked by **|**, similarities by **:**. Every tenth amino acid is in bold typeface. Small typeface for Phe 344 and Asn 399 indicate residues excluded from the modelling.

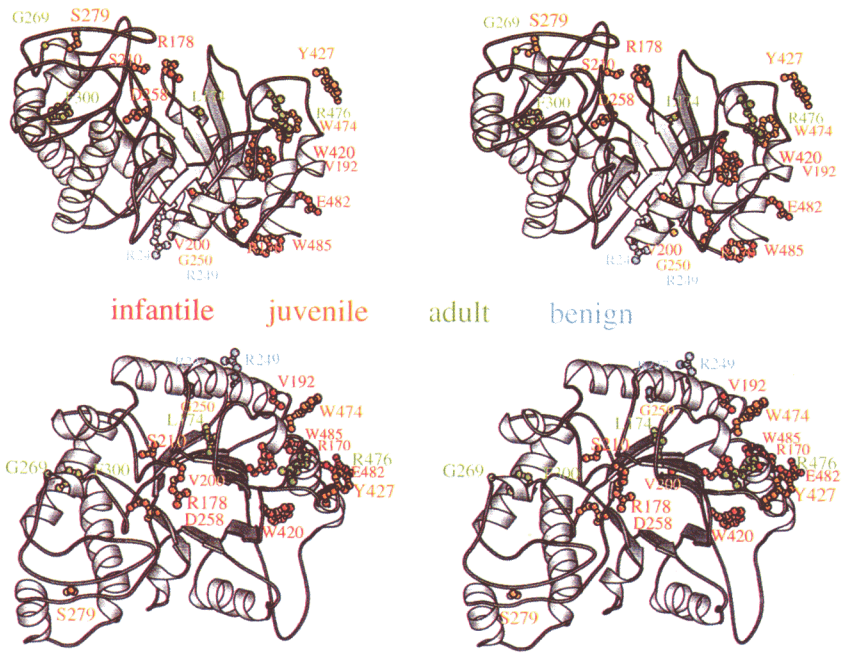


Fig. 6 Predicted structure of the $(\beta\alpha)_8$-barrel of hexosaminidase showing pathogenic mutations colour coded by phenotypical classification (see text and Table 3): infantile mutations are in red, juvenile mutations are orange, adult mutations green, and benign mutations blue. The lower view shows the barrel from the top, the upper view is turned –90° around the *x*-axis. Note that the $(\beta\alpha)_8$-barrel domain is more compact than that in chitobiase.

chitobiases with 42–54% overall sequence identity (this study and refs 14,29,30); the second group contains two mammalian hexosaminidase α-chains (HEXA) with 84% overall identity[31,32]; the third group contains three mammalian hexosaminidase β-chains (HEXB) with up to 81 % overall identity[33–35]. HEXA and HEXB are closely related[36]. There are four other eukaryotic hexosaminidases making up the intermediate branch, which are excluded from the present analysis as they are less related.

Homology modelling of hexosaminidases is valuable because defects in these enzymes lead to various disease phenotypes[12,13]. We confine the analysis here to the relationships between our bacterial chitobiases and the hexosaminidase α-chains. The starting data were the structure of chitobiase and the multiply aligned sequences of the hexosaminidases. Alignments of chitobiase with hexosaminidase sequences using ClustalW[37] showed identities of about 26% for the catalytic domain. Homology modelling was restricted to the catalytic domain only.

With 26% sequence identity homology modelling is difficult but nevertheless possible[38,39]. Similar hydrophobic cores and active sites are expected with the differences restricted to the surface loops. We used a classical multiple sequence alignment with Clustal W optimized by PHD[40]. The resulting sequence alignment (Fig. 5) was the basis for homology modelling using WHATIF[41–43] (Methods) and the model is shown in Fig. 6. Several observations support the idea that the alignment is correct. First, in the homology model all active-site residues could easily be aligned with those in the chitobiase structure. Second, in the model there is a well defined hydrophobic core with no van der Waals clashes between residues; there are also about the same number of small cavities as found in the template structure. All parameters from the BIOTECH validation[41] are similar for the template structure and the proposed model (data not shown). Further support for the alignment comes from the location of the deletions and insertions, mainly regions present in chitobiase but not in hexosaminidase. Examples are residues 400–418, 461–477, 544–569 and 675–687, which lie in loops between the secondary structural elements of chitobiase. This is in keeping with the observation that the loops in different but homologous $(\beta\alpha)_8$-barrels can vary greatly in length.

Additional support for the model comes from the observation of pairs or groups of residue substitutions which complement each other in the three-dimensional structures of the two enzymes (Table 2). These pairs of substitutions are assumed to have occurred during evolution to complement one another in maintaining the structural integrity of the protein. Although it is certain that the HEXA model contains many errors, the high degree of conservation of residues in and around the active site means that the model is much less prone to errors in the active site pocket.

Tay–Sachs and Sandhoff disease

Deficiency in hexosaminidases cause lysosomal storage diseases[12]. There are two human hexosaminidase genes, encoding proteins that form dimers with different substrate specificity[42]. The homodimer S (αα) has no measurable activity while the homodimer B (ββ) is active against certain oligosaccharides. The heterodimer A (αβ) has full activity against G_{M2} type gangliosides, accumulation of which can cause blindness, paralysis, dementia and death. Tay–Sachs disease (TSD) is caused by deficient or defective hexosaminidase α-subunit. 97% of TSD carriers are positive in one allele for either an insertion of four base pairs introducing an early stop codon or a mutation causing mis-splicing[13]. In pathological cases, one of these gene deficiencies is frequently observed in combination with a missense mutation in the second allele. TSD is termed the 'B variant' if the α-subunit is absent or present in lower than normal amounts, as the homodimer B (ββ) is present at normal levels but those of the homodimer S (αα) and the heterodimer A (αβ) are lowered. The 'B1 variant', with normal levels of all three dimers, is caused by a present but inactive α-subunit. Sandhoff disease (SD), observed in several geographic or demographic isolates, involves deficiency or defect in the β-subunit. It is termed the '0 vari-

ant' as there is only the inactive hexosaminidase S ($\alpha\alpha$) present at normal levels. The absence of hexosaminidase A (resulting in a lack of $\alpha\beta$) causes G_{M2} accumulation, lack of hexosaminidase B (resulting in a lack of $\beta\beta$) in addition leads to accumulation of oligosaccharides (for example, from glycoproteins). This explains the different phenotype to TSD. Onset and severity of the clinical course of TSD and SD correlate with enzyme activity. Complete loss of function leads to the infantile form. The juvenile form has residual activity while adult and benign forms retain substantial activity. Wide variability in late onset is observed. Various factors may cause pathological states: unsuccessful folding, insolubility, failure to transport to the lysosome, instability within the lysosome, failure to form the heterodimer, failure to bind activator protein or loss of enzyme activity.

Table 2 Alternate complementary residues in the structure of chitobiase (Chb) and proposed structure of hexosaminidase A (HexA)

Enzyme	Residues	Contacts/comments
Chb	Ile 384-Tyr 429-Phe 374-Val 357	
HexA	Tyr 213-Val 239-Trp 203-Ile 186	Altered hydrophobic packing
Chb	Ala 456-Trp 382	
HexA	Trp 266-Phe 211	Altered hydrophobic packing
Chb	Ile 384-Met 523	
HexA	Tyr 213-Val 311	Space compensation
Chb	Ile 386-Val 516	
HexA	Ser 215-Phe 304	Space compensation
Chb	Gly 388-Lys 515	Surface or interface maintenance
HexA	Thr 217-Thr 303	and space compensation
Chb	Gly 397-Phe 423	
HexA	Ser 226-Ile 233	Space compensation
Chb	Ile 447-Ile 520	
HexA	Phe 257-Val 308	Altered hydrophobic packing
Chb	Met 449-Val 513	
HexA	Thr 269-Met 301	Methionine interchange
Chb	Asp 483-Val 577	Swapped charged groups, surface
HexA	Leu 274-Glu 334	maintenance and space compensation
Chb	Cys 578-Cys 505	
HexA	Ile 335-Ser 293	Compensation of Cys bridge
Chb	Arg 511-Ser 508	Change in surface character
HexA	Glu 299-Asn 296	and space compensation
Chb	Ile 517-Trp 534	
HexA	Phe 305-Leu 317	Change in hydrophobic packing
Chb	Ile 384-Met 523-Tyr 429	
HexA	Tyr 213-Val 311-Val 239	Tyrosine interchange
Chb	Lys 570-Glu 591	Swap charge on surface, change
HexA	Thr 327-Lys 349	of character, space compensation
Chb	Phe 536-Gly 598	
HexA	Ile 319-Ile 356	Change of hydrophobic packing
Chb	Glu 659-Arg 633	
HexA	Arg 413-Asp 386	Swap charges on surface

A set of substitutions between the known three-dimensional structure of chitobiase (Chb) and that modelled for hexosaminidase A (Hex A). The sets of pairs of residues in the two enzymes are complementrary in the sense that they are mutually compensating and thus allow the integrity of the protein fold to be maintained. For instance the pair Phe 536-Gly 598 in Chb has one large and one small side chain: the equialent pair in HexA is Ile 413-Gly 598 with two medium sized hydrophobic side chains. Other parings involve for example opposite charges on residue pairs as well as merely size compensation. The indicated sets of residues appear to have evolved so that they make up equivalent structural units in the two molecules.

In an effort to understand the various manifestations of the disease, we mapped the known mutations onto the homology-modelled hexosaminidase structure (Fig. 6, Table 3). Mutations leading to the severe, infantile disease lie in two clusters. The first cluster is around the active site and causes the B1 phenotype. These mutations include Arg178Cys/His/Leu (refs 45–47), Ser210Phe (ref. 48), Asp258His (ref. 49) and Trp420Cys (ref. 50). (Residue numbers quoted in the text refer to the α-chain (Table 3); equivalent residues from chitobiases are in parentheses throughout the text.) Asp258His (Asp 448) causes a pH optimum shift[49]; this residue is involved in substrate binding (Figs 2*a*, 3). Ser210Phe (Gly 381) resides in a loop with three key functional residues (Asp 378, Asp 379, Glu 380) (Figs 2*d*, 3). The mutation may disrupt the three-dimensional structure of the loop. Asp258His and Ser210Phe probably introduce steric hindrance either within the protein fold or in enzyme–substrate interactions. Trp420Cys has been proposed to introduce a new Cys-bridge[50] but may well introduce a conformational change close to Tyr 421 (Tyr 669) which is involved in substrate binding (Fig. 2*f*). The involvement of the Arg 178 mutations in this first, severe cluster is notable. Assuming the homology model based on the chitobiase structure is correct, Arg 178 (Arg 349) is directly involved in substrate binding (Figs 2*d*, 3). The importance of Arg 178 of hexosaminidase is underscored by experiments showing that when mutations in this α-subunit were duplicated in the β-subunit, inactivity resulted[51]. In this light, the so-called 'active arginine' hypothesis postulating a direct charge-transfer role in catalysis for this residue[52] seems incorrect.

The second cluster of infantile mutations comprises Arg170Gln/Trp (refs 49,53), Glu482Lys (ref. 54), Trp485Arg (ref. 55) and the double mutation Val192Leu / Val200Met (ref. 56). Some of these mutations are described as B1 phenotypes. Nevertheless, our model places these residues far from the active site on the N-terminal

Table 3 Point mutations in hexosaminidases grouped according to onset of the disease

type	Hex A	Hex B	Chb	Position in Chb	Phenotype	Class	Ref.
Infantile	R170Q/W	(R203)	R341	begin β1	unstable, slow maturation	see text	49,53
	R178C/H/L	(R211)	R349	loop 1	no activity	B1	45–47
	V192L	(A225)	Q363	centre α1	<2 % activity	see text	56
	+V200M	(I233)	K371	begin β2	neutral exchange	see text	56
	S210F	(S243)	G381	begin loop 2	no activity	B1	48
	D258H	(D290)	D448	end β3	pH optim. shift	B1	49
	ΔF305	(F337)	I517	center of α3	inactive precursor	see text	48, 77
	W420C	(W449)	V668	begin loop 7	no activity	B1	50
	E482K	(E511)	E760	end α8	remains in ER, < 1 % activity	sol	54
	W485R	(W514)	W763	end α8	chinese infantile TSD	see text	55
Juvenile	G250D	(G382)	Q440	begin β3	2–3 % activity	unknown	62
	S279P	(S311)	T487	loop 3	lower RNA level	unknown	77
	(Y427)	Y456S	E675	loop 7	degradation	thl	78
	W474C	(W503)	(F752)	centre of α8	12 %, encephalopathic	thl	61
	R499C/H	(R528)	n.m.	not modelled	remains in ER, 3% activity	sol	79,80
	R504C/H	(R533)	n.m.	not modelled	inh. dimeriz.	dim	48, 80, 81
Adult	(L174)	I207V	L345	centre of β1	reduced activity	thl	78
	G269S	(G301)	S459	loop 3, long helix	12 %	dim	62,63,65
	F300L	(F332)	F512	begin α3	chronic form	thl	82
	(R476)	R505Q	R754	center of α8	thermolabile	thl	60
Benign	R247W	(R279)	Q437	end α2	40 %, pseudodeficient	benign	58,62
	R249W	(R281)	R439	end α2	40 %, pseudodeficient	benign	59,62

Homologous amino acid positions are shown for hexosaminidase α-subunit (HexA), β-subunit (HexB) and for chitobiase (Chb). Parentheses around residue numbers indicate equivalent positions for the known mutated residues which are given without parentheses. For the position of the residues, compare with Figs 1 and 6. The observed cellular activity (from the original publications) is listed as phenotype, a separate column gives an interpretation on a structural basis classifying the mutations as: benign—no phenotype but reduced activity, dim—inhibited homo/heterodimer formation, thl—thermolabile enzyme with reduced half life time, sol—enzyme with altered solubility, and B1—without activity, but normal as characterized with antibodies against the enzyme. Not modelled (n.m.) are two mutations that fall just outside the αβ-fold where no reliable prediction could be made.

end of the barrel suggesting that other effects may cause the disease. Arg170Gln/Trp leads to unstable enzyme with a low rate of maturation[45]. In the double mutation Val192Leu/Val200Met only Val192Leu is responsible for the B1 phenotype[48,57]. From our model it remains unclear how such a small change causes the disease. Glu482Lys leads to insoluble protein which is trapped unprocessed in the endoplasmic reticulum[46]. The clustering of Arg 170, Val 192, Glu 482 and Trp 485 suggests the possibility that they are part of an interface either for domain–domain or subunit–subunit interaction.

The other mutations (Table 3 and Fig. 6) are less severe in phenotype. Most of these follow a general scheme: benign and adult mutations are often on the protein surface where a change does not exert a significant effect on protein stability; more severe, juvenile, mutations are closer to the protein core. They all lie well away from the active site. This is exemplified by the benign mutations Arg247Trp (ref. 58) and Arg249Trp (ref. 59) the adult mutation Arg505Gln (ref. 60) and the juvenile mutation Trp474Cys (ref. 61). Arg247Trp/ Arg249Trp, although apparently large changes, are exposed to solvent and retain significant activity[62]. Trp474Cys is buried, alters the hydrophobic packing and is important for protein stability. Arg505Gln (observed in SD for the β-chain, equivalent to Arg 476 in the α-chain, Fig. 6) disrupts important salt bridges and induces thermolability[52]. Buried mutations may change thermostability, and mutations on the surface may alter conditions for domain interaction, protein–protein interaction or solubility with consequent transport defects.

There are mutations which do not follow these general principles. The adult mutation Gly269Ser which affects dimerization is a good example[63–65]. Chitobiase has a serine at this position, which is in the middle of the long helix in loop 3 (Fig. 1). In the hexosaminidase model, this helix terminates after one turn with Gly 269 into a loop structure: in comparison with chitobiase the remaining helix and a second short helix are deleted. The mutation from Gly 269 to serine will probably not allow the altered enzyme to terminate the helix efficiently in a sharp turn and the altered loop structure may then affect dimerization. These results, based on homology modelling, have to be considered carefully in the light of functional and environmental differences to periplasmic bacterial chitobiase, for example the disulphide bridge pattern. Nevertheless, the locations of the mutations deduced from this model cluster within the proposed structure reasonably well according to the severity of the phenotype. It will be interesting to see if analogous mutations in chitobiase will lead to similar defects.

Methods

Sequencing and protein purification. Genomic DNA from *S. marcescens* cloned into pEMBL18 (ref. 66) was used to constitutively express chitobiase in *E. coli*. The deduced 98,500 M_r protein has 885 amino acids. Proteolytic cleavage of its 27 amino acid leader peptide was verified by protein

sequencing. Recombinant chitobiase from the bacterial periplasm was initially fractionated with ammonium sulphate (55% to 85%); the protein was then further purified by anionic ion exchange chromatography (S-Sepharose; Pharmacia), eluting at pH 6.8 at about 120 to 250 mM NaCl. A second purification step with Mono-S FPLC (Pharmacia) at pH 6.4 and a concentration step on the same column yielded the pure protein at a concentration of up to 30 mg ml^{-1}.

Structure determination. Crystals grew in the orthorhombic space group $P2_12_12$ within four weeks from 61.5% ammonium sulphate and 1.5% isopropanol in 100 mM cacodylate buffer, pH 5.6, at room temperature. Unit cell dimensions are (Å) *a*=110.7, *b*=99.9, *c*=87.7. Individual data sets were collected on single crystals using a Marresearch imaging plate scanner at the EMBL beam lines X31 (native, derivatives) and BW7B (diNAG). Reflections were integrated and scaled using DENZO and SCALEPACK[67]. Multiple isomorphous replacement (MIR) phasing was carried out using gold and platinum derivatives. All derivatives and complexes were prepared by soaking crystals in appropriate solutions: Gold 12 h in 10 mM $KAuCl_4$; Platinum 4 h in 20 mM Pt(II)-(2,2'-6',2''-tetrapyridine)-chloride-dihydrate; diNAG 290 s in 100 mg ml-1 diNAG (2-acetamido-2-deoxy-4-O-(2-acetamido-2-deoxy-b-d-glucanopyranosyl)-d-glucopyranose, trivial name N,N'-diacetylchitobiose or chitobiose). The PHASES package[68] was used for heavy atom refinement, solvent flattening and map skeletonization. The MIR phase set had a mean figure of merit of 0.69. The model was built into solvent-flattened MIR density using O[69]. Prior to refinement, 10% of the data were randomly flagged for cross validation using R_{free}[70]. Restrained Hendrickson–Konnert refinement[71] with the program PROLSQ from the CCP4 suite[72] was used. The ARP program[73] was used to locate automatically water molecules. Refinement converged to an R_{factor} of 13.9 % with R_{free} 19.6 %. The R_{factor} given in Table 1 was calculated including all measured data. Comparisons against the database of known structures were carried out with DALI[74].

diNAG complex. Soaking was performed by adjusting the crystallisation buffer to the desired diNAG concentration. A crystal was exposed to this solution for about five minutes. Initially difference maps were calculated against structure factors from the final protein model omitting the waters in the active site pocket. Refinement was carried out using an updated protein dictionary with the small molecule structure of diNAG[23]. The coordinates and structure factors have been deposited with the Brookhaven Protein Data Bank. Coordinates—native chitobiase: 1QBA; chitobiase–diNAG: 1QBB; homology model of human HEXA: 1QBC; homology model of human HEXB: 1QBD. Structure factors—native chitobiase: R1QBASF; chitobiase–diNAG: R1QBBSF.

Hexosaminidase alignment and modelling. As a first step, the multiple alignments of the hexosaminidases were input to PHD to predict the secondary structural elements. Eight β-strands and six α-helices were predicted for hexosaminidases, in the correct order to compose most of the secondary structural elements of $(\beta\alpha)_8$-barrel. The sequence alignment of chitobiase and hexosaminidase were then adjusted manually to optimize the agreement between the predicted secondary structure elements of hexosaminidase and the known secondary structure elements in chitobiase. This in principle optimizes the use of the available three-dimensional information for chitobiase, giving an improved alignment over primary structural data alone. The resulting sequence alignment was the basis for homology modelling using WHATIF[41]. A model for the backbone structure of hexosaminidase was built from that of chitobiase. The modelling protocol described in ref. 42 was used with the default parameters. Cα positions were copied from chitobiase for aligned segments. Ala, Gly, and Pro were placed directly using chitobiase coordinates. All other residues were initially reduced to Ala and mutated one by one to the final residue starting with residues having the fewest rotamers available[43] in the given structural environment. Alignments were adjusted by inspecting the models. The underlying alignment was corrected and a new model built. This process was iterated until the model satisfied both the geometric criteria of WHATIF[41] criteria validation suite and our knowledge about conserved residues (Fig. 6). The final model was energy minimized by GROMOS[75] and regularized by WHAT-IF[41].

Received 4 January; accepted 18 April 1996.

1. Henrissat, B. & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788 (1993).
2. Davies, G.J. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
3. Chet, I., Barak, Z. & Oppenheim, A.B. Genetic engineering of microorganisms for improved biocontrol activity. in *Biotechnology in Plant Disease Control* (ed. I. Chet) 211–235 (Wiley Liss, Inc. 1993).
4. ZoBell, C.E. & Rittenberg, S.C. The occurence and characteristics of chitinoclastic bacteria in the sea. *J. Bacteriol.* **35**, 275–287 (1937).
5. Perrakis, A. *et al.* Crystal structure of a bacterial chitinase at 2.3 Å resolution. *Structure* **2**, 1169–1180 (1994).
6. Roey, P.V., Rao, V., Plummer Jr., T.H. & Tarentino, A.L. Crystal Structure of Endo-β-N-acetylglucosaminidase F1, an α/β-Barrel Enzyme Adapted for a Complex Substrate. *Biochem.* **33**, 13989–13996 (1994).
7. Rao, V., Guan, C. & Roey, P.V. Crystal structure of endo-β-N-acetylglucosaminidase H at 1.9 Å resolution: active-site geometry and substrate recognition. *Structure* **3**, 449–457 (1995).
8. Terwisscha van Scheltinga, A.C., Kalk, K.H., Beintema, J.J. & Dijkstra, B.W. Crystal structure of hevamine, a plant defence protein with chitinase and lysozyme activity. *Structure* **2**, 1181–1189 (1994).
9. Hart, P.J., Pfluger, H.D., Monzingo, A.F., Hollis, T. & Robertus, J.D. The Refined Crystal Structure of the Endochitinase from Hordeum vulgare L. Seeds at 1.8 Å Resolution. *J. Mol. Biol.* **248**, 402–413 (1995).
10. Oppenheim, A.B. & Chet, I. Cloned chitinases in fungal plant-pathogen control strategies. *Trends. Biotech.* **10**, 392–394 (1992).
11. Roberts, R.L. & Cabib, E. Serratia marcescens Chitinase: One-Step Purification and Use for the Determination of Chitin. *Analyt. Biochem.* **127**, 402–412 (1982).
12. Sandhoff, K., Conzelmann, E., Neufeld, E.F., Kaback, M.M. & Suzuki, K. *The G_{M2} ganglosidoses. in The Metabolic Base of Inherited Disease 6th ed.* (ed. Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D.) 1807–1842 (McGraw-Hill, New York, 1988).
13. Mahuran, D.J. The biochemistry of HEXA and HEXB gene mutations causing G_{M2} gangliosidosis. *Biochim. Biophys. Acta.* **1096**, 87–94 (1991).
14. Tews, I., Vincentelli, R. & Vorgias, C.E. *N*-acetylglucosaminidase (chitobiase) from *Serratia marcescens*: Gene sequence, and protein production and purification in *Escherichia coli*. *Gene* **170**, 63–67 (1996).
15. Xu, G.-Y. *et al.* Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochem.* **34**, 6993–7009 (1995).
16. Baumann, U. Crystal Structure of the 50 kDa Metallo protease from *Serratia marcescens*. *J. Mol. Biol.* **242**, 244–251 (1994).
17. Banner, D.W. *et al.* Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution. *Nature* **255**, 609–614 (1975).
18. De Vos, A.M., Ultsch, M. & Kossiakoff, A.A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312 (1992).
19. Koshland, D.E. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416–436 (1953).
20. Paulsen, H. Fortschritte bei der selektiven chemischen Synthese komplexer Oligosaccharide Angew. *Chem. Int. Ed.. Engl.* **21**, 155–173 (1982).
21. Sinnot, M.L. Stereochemistry and the mechanisms of enzymatic glycosyl transfer. *Chem Rev.* **90**, 1171–1202 (1990).
22. McCarter, J.D. & Withers, S. Mechanisms of enzymatic glycoside hydrolysis. *Curr. Op. Struct. Biol.* **4**, 885–892 (1994).
23. Mo, F. & Jensen, L.H. The crystal structure of β-(1→4) linked disaccharide, α-N,N'-diacetylchitobiose monohydrate. *Acta Cryst.* B**34**, 1562–1569 (1978).
24. Lai, E.C.K. & Withers, S.G. Stereochemistry and kinetics of the hydration of 2-acetamido-D-glucal by β-*N*-acetylhexosaminidases. *Biochem.* **33**, 14743–14749 (1994).
25. Thunissen, A-M.W.H. *et al.* Doughnut-shaped structure of a bacterial muraminidase revealed by X-ray crystallography. *Nature* **367**, 750–753 (1994).
26. Weaver, L.H., Grütter, M.G. & Matthews, B.W. The refined structures of

Acknowledgements

We thank R. Vincentelli for technical assistance, G. Vriend for help with homology modelling, L. Holm for database searches, B. Rost for help with secondary structure prediction, S. Armand and B. Henrissat, Centre de Recherches sur les Macromolécules Végétales, CNRS, Grenoble, France for supplying biochemical data in advance of publication, and C. Wiesmann, University of Freiberg, Germany for the suggestion of the short soaking time. The work was partly supported by an EU grant to C.E.V. and an EU Human Capital Mobility Program Institutional fellowship to A.P.

goose lysozyme and its complex with a bound trisaccharide show that the "goose-type" lysozyme lack a catalytic aspartate residue. *J. Mol. Biol.* **245**, 54–68 (1995).
27. Terwisscha van Scheltinga, A.C. *et al.* Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin. Evidence for substrate assisted catalysis. *Biochem.* **34**, 15619–115623 (1995).
28. Lowe, G. & Sheppard, G. Acetamido-group participation in lysozyme catalysis. *J. Chem. Soc. Chem. Commun.* 529–530 (1968).
29. Soto-Gil, R.W. & Zyskind, J.W. N,N'-Diacetylchitobiase of *Vibrio harveyi*. *J. Biol. Chem.* **264**, 14778–14783 (1989).
30. Somerville, C.C. & Colwell, R.R. Sequence analysis of the b-N-acetylhexosaminidase gene from *Vibrio vulnificus*: Evidence for a common evolutionary origin of hexosaminidases. *Proc. Natl. Acad. Sci. USA* **90**, 6751–6755 (1993).
31. Beccari, T., Hoade, J., Orlacchio, A. & Stirling, J.L. Cloning and sequence analysis of a cDNA encoding the α-subunit of mouse β-N-acetylhexosaminidase and comparison with the human enzyme. *Biochem. J.* **285**, 593–596 (1992).
32. Myerowitz, R., Piekarz, R., Neufeld, E.F., Shows, T.B. & Suzuki, K. human β-hexosaminidase a chain: Coding sequence and homology with the β chain. *Proc. Natl. Acad. Sci. USA* **82**, 7830–7834 (1985).
33. Bapat, B., Ethier, M., Neote, K., Mahuran, D. & Gravel, R.A. Cloning and sequence analysis of the cDNA encoding the β-subunit of mouse β-hexosaminidase. *FEBS Lett.* **237**, 191–195 (1988).
34. Muldoon, L.L., Neuwelt, E.A., Pagel, M.A. & Weiss, D.L. Characterization of the molecular defect in feline model for type II GM2 gangliosidoesis (Sandhoff disease) *Am. J. Pathol.* **144**, 1109–1118 (1994).
35. Neote, K. *et al.* Characterization of the human HEXB gene encoding lysosomal β-hexosaminidase. *Genomics* **3**, 279–286 (1988).
36. Korneluk, R.G. *et al.* Isolation of cDNA clones coding for the α-subunit of human β-hexosaminidase. *J. Biol. Chem.* **261**, 8407–8413 (1986).
37. Higgins, D.G., Bleasby, A.J. & Fuchs, R. CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* **8**, 189–191 (1992).
38. Sander, C. & Schneider, R. Database of homogy-derived structures and the structurally meaning of sequence alignment. *Proteins* **9**, 56–68 (1991).
39. Chothia C. & Lesk, A.M. The relation between the divergence of sequence and structure in protein. *EMBO J.* **5**, 823–826 (1986).
40. Rost, B. & Sander, C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**, 55–72 (1994).
41. Vriend, G. WHAT-IF: A molecular modelling and drug design program. *J. Mol. Graph.* **8**, 52–56 (1990).
42. Chinea, G., Padron, G., Hooft, R.W.W., Sander, C. & Vriend, G. The use of protein-specific rotamers in model building by homology. *Proteins* **23**, 415–421 (1995).
43. de Fillippis, V., Sander, C. & Vriend, G. Predicting local structural changes that result from point mutations. *Prot. Engng.* **7**, 1203–1208 (1994).
44. Kytzia, H.-J. & Sandhoff, K. Evidence for two different active sites on human β-hexosaminidase A. *J. Biol. Chem.* **260**, 7568–7572 (1985).
45. Tanaka, A. *et al.* GM2-gangliosidosis B1 variant: analysis of β-hexosaminidase a gene abnormalities in seven patients. *Am. J. Hum. Genet.* **46**, 329–339 (1990).
46. Ohno, K. & Suzuki, K. Mutation in GM2 gangliosidosis B1 variant. *J. Neurochem.* **50**, 316–318 (1988).
47. Triggs-Raine, B.L., Akerman, B.R., Clarke, J.T.R. & Gravel, R.A. Sequence of DNA flanking the exons of the HEXA gene, and identification of mutations in Tay–Sachs Disease. *Am. J. Hum. Genet.* **49**, 1041–1054 (1991).
48. Akli, S., Chelly, J., Lacorte, J.-M., Poenaru, L. & Kahn, A. Seven novel Tay–Sachs mutations detected by chemical mismatch cleavage of PCR-amplified cDNA fragments. *Genomics* **11**, 124–134 (1991).
49. Fernandes, M. *et al.* A new Tay–Sachs disease B1 allele in exon 7 in two compound heterozygotes each with a second novel mutation. *Hum. Mol. Genet.* **1**, 759–761 (1992).
50. Tanaka, A., Punnett, H.H. & Suzuki, K. A new point mutation in the β-hexosaminidase a subunit gene responsible for infantile Tay–Sachs Disease in a non-Jewish caucasian patient (a Kpn Mutant). *Am. J. Hum. Genet.* **47**, 567–574 (1990).
51. Brown, C.A., Neote, K., Leung, A., Gravel, R.A. & Mahuran, D.J. Introduction of the a subunit mutation associated with the B1 variant of Tay–Sachs Disease into the β subunit produces a β-hexosaminidase B without catalytic activity. *J. Biol. Chem.* **264**, 21705–21710 (1989).
52. Brown, C.A. & Mahuran, D.J. Active arginine residues in β-hexosaminidases. *J. Biol. Chem.* **266**, 15855–15862 (1991).
53. Nakano, T. *et al.* A new point mutation within exon 5 of β-hexosaminidase a gene in a Japanese infant with Tay–Sachs Disease. *Ann. Neurol.* **27**, 465–473 (1990).
54. Nakano, T., Muscillo, M., Ohno, K., Hoffman, A.J. & Suzuki, K. A point mutatation in the coding sequence of the β-hexosaminidase a gene results in the enzyme protein in an unusual G_{M2}-gangliosidosis variant. *J. Neurochem.* **51**, 984–987 (1988).
55. Akalin, N. *et al.* Tay–Sachs Disease in China: two new mutations and a "Macro Polo" allele. *Am. J. Hum. Genet.* **49**, A2246 (1991).
56. Ainsworth, P.J. & Coulter-Mackie, M.B. A double mutation in exon 6 of the β-hexosaminidase α subunit in a patient with the B1 variant of Tay–Sachs Disease. *Am. J. Hum. Genet.* **51**, 802–809 (1992).
57. Coulter-Mackie, M.B. Molecular characterization of both alleles in an unusual Tay–Sachs Disease B1 variant. *Am. J. Hum. Genet.* **54**, 1126–1127 (1994).
58. Triggs-Raine, B.L. *et al.* A pseudodeficiency allele common in non-Jewish Tay–Sachs carriers: implications for carrier screening. *Am. J. Hum. Genet.* **51**, 793–801 (1992).
59. Cao, Z. *et al.* A second mutation associated with apparent β-hexosaminidase A pseudodeficiency: identification and frequency estimation. *Am. J. Hum. Genet.* **53**, 1198–1205 (1993).
60. Bolhuis, P.A., Ponne, N.J., Bikker, H., Baas, F. & deJong, V.J.M.B. Molecular basis of an adult form of Sandhoff disease: substitution of glutamine for arginine at position 505 of the β-chain of β-hexosaminidase results in a labile enzyme. *Biochim. Biophys. Acta.* **1182**, 142–146 (1993).
61. Petroulakis, E., Cao, Z., Salo, T., Clarke, J. & Triggs-Raine, B. A Trp474Cys mutation in the alpha-subunit of beta-hexosaminidase causes the subacute encephalopathic form of GM2 gangliosidosis, Type 1. *Am. J. Hum. Genet.* **55**, A2129 (1994).
62. Cao, Z., Petroulakis, E., Salo, T. & Triggs-Raine, B. Expression of the HEXA mutations Arg247Trp and Arg249Trp, associated with beta-hexosaminidase A pseudodeficiency. *Am. J. Hum. Genet.* **55**, A1251 (1994).
63. Navon, R. & Proia, R.L. The mutations in Ashkenazi Jews with adult GM2 gangliosidosis, the adult form of Tay–Sachs Disease. *Science* **243**, 1471–1474 (1989).
64. Paw, B.H., Kaback, M.M. & Neufeld, E.F. Molecular basis of adult-onset and chronic G_{M2} gangliosidosis of Ashkenazi Jewish origin: substitution of serine for glycine at position 269 of the α-subunit of β-hexosaminidase. *Proc. Natl. Acad. Sci. USA* **86**, 2413–2417 (1989).
65. Brown, C.A. & Mahuran, D.J. β-hexosaminidase Isozymes from cells cotransfected with α and β cDNA constructs: analysis of the α-subunit missense mutation associated with the adult form of Tay–Sachs Disease. *Am. J. Hum. Genet.* **53**, 497–508 (1993).
66. Kless, H., Sitrit, Y., Chet, I. & Oppenheim, A.B. Cloning of the gene coding for chitobiase of *Serratia marcescens*. *Mol. Gen. Genet.* **217**, 471–473 (1989).
67. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meths. Enzymol.* **276** (in the press).
68. Furey, W. & Swaminathan, S. PHASES - a program package for the processing and analysis of diffraction data from macromolecules. *American Crystallographic Association Meeting Abstracts* **18**, 73 (1990).
69. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr.* A**47**, 110–119 (1991).
70. Brünger, A.T. Assessment of phase accuracy by cross validation: the free R value. Methods and application. *Acta Crystallogr.* D**49**, 24–36 (1993).
71. Konnert, J.H. & Hendrickson, W.A. A restrained-parameter thermal-factor refinement procedure. *Acta Crystallogr.* A**36**, 344–350 (1980).
72. Collaborative Computer Project, Number 4 (CCP4) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.* D**50**, 760–763 (1994).
73. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta Crystallogr.* D**49**, 129–147 (1993).
74. Holm, L. & Sander, C. Protein Structure Comparison by Alignment of Distance Matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
75. van Gunsteren, W.F. & Berendsen, H.J. GROMOS. BIOMOS, Biomolecular Software, University of Groningen, the Netherlands .
76. Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plot of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
77. Navon, R. & Proia, R.L. Tay–Sachs Disease in Moroccan Jews: deletion of a phenylalanine in the α-subunit of β-hexosaminidase. *Am. J. Hum. Genet.* **48**, 412–419 (1991).
78. Banerjee, P. *et al.* Molecular basis of an adult form of β-hexosaminidase B deficiency with motor neuron disease. *Biochem. Biophys. Res. Comm.* **181**, 108–115 (1991).
79. Mules, E.H., Hayflick, S., Miller, C.S., Reynolds, L.W. & Thomas, G.H. Six novel deleterious and three neutral mutations in the gene encoding the α-subunit of the hexosaminidase A in non-Jewish individuals. *Am. J. Hum. Genet.* **50**, 834–841 (1992).
80. Paw, B.H. *et al.* Juvenile GM2 gangliosidosis caused by substitution of histidine for arginine at position 499 or 504 of the α-subunit of β-hexosaminidase. *J. Biol. Chem.* **265**, 9452–9457 (1990).
81. Paw, B.H., Wood, L.C. & Neufeld, E.F. A third mutation at the CpG dinucleotide of codon 504 and a silent mutation at codon 506 of the HEXA Gene. *Am. J. Hum. Genet.* **48**, 1139–1146 (1991).
82. Fernandes, M., Boulay, B., Hechtmann, P., Kaplan, F. & Strasberg, P. Five novel HEXA mutations in non-Jewish Tay–Sachs disease (TSD) patients. *Am. J. Hum. Genet.* **51**, A656 (1992).