THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 19, pp. 14300–14308, May 11, 2007

Structural and Mechanistic Analyses of *endo*-Glycoceramidase II, a Membrane-associated Family 5 Glycosidase in the Apo and G_{M3} Ganglioside-bound Forms*

Received for publication, December 14, 2006, and in revised form, February 8, 2007 Published, JBC Papers in Press, February 28, 2007, DOI 10.1074/jbc.M611455200

Matthew E. C. Caines[‡], Mark D. Vaughan[§], Chris A. Tarling[§], Susan M. Hancock[§], R. Antony J. Warren[¶], Stephen G. Withers[‡§], and Natalie C. J. Strynadka[‡1]

From the Departments of [‡]Biochemistry and Molecular Biology, [§]Chemistry, and [¶]Microbiology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

***endo*-Glycoceramidase, a membrane-associated family 5 glycosidase, deviates from the typical polysaccharide substrate specificity of other soluble members of the family, preferentially hydrolyzing glycosidic linkages between the oligosaccharide and ceramide moieties of gangliosides. Here we report the first x-ray crystal structures of an *endo*-glycoceramidase from *Rhodococcus* sp., in the apo form, in complex with the ganglioside G_{M3} (Svennerholm ganglioside nomenclature (Svennerholm, L. (1964) *J. Lipid Res.* 5, 145–155)), and trapped as a glycosyl-enzyme intermediate. These snapshots provide the first molecular insight into enzyme recognition and association with gangliosides, revealing the structural adaptations necessary for glycosidase-catalyzed hydrolysis and detailing a novel ganglioside binding topology. Consistent with the chemical duality of the substrate, the active site of *endo*-glycoceramidase is split into a wide, polar cavity to bind the polyhydroxylated oligosaccharide moiety and a narrow, hydrophobic tunnel to bind the ceramide lipid chains. The specific interactions with the ceramide polar head group manifest a surprising aglycone specificity, an observation substantiated by our kinetic analyses. Collectively, the reported structural and kinetic data provide insight toward rational redesign of the synthetic glycosynthase mutant of *endo*-glycoceramidase to enable facile synthesis of nonnatural, therapeutically useful gangliosides.**

Gangliosides are prominent components of mammalian cell membranes that are particularly prevalent in neuronal membranes (1). They are implicated in numerous signaling roles, with involvement in the cell cycle, differentiation, communication, recognition, and apoptosis (2). In mammalian systems, gangliosides are synthesized by the sequential, glycosyltransferase-catalyzed, addition of monosaccharides to a ceramide (1). In a similar stepwise manner, they are catabolized by *exo*-glycosidases that cleave monosaccharides from the nonreducing oligosaccharide end. In contrast, *endo*-glycosidases that directly hydrolyze the glycosidic linkage between the ceramide and oligosaccharide (Fig. 1*a*) have been discovered in non-mammalian cells (3–9), where they presumably play a nutritional role for the organism. One such enzyme, *endo*-glycoceramidase II (EGC)[2] from *Rhodococcus*, is a family 5 glycosidase (www.cazy.org) that hydrolyzes gangliosides with a net retention of stereochemistry at the anomeric center via the double displacement mechanism common to other retaining glycosidases (Fig. 1*b*) (10).

EGC is noticeably distinct in function from other family 5 glycosidases, including the prototypical β-(1,4)-glucanases (cellulases), β-(1,4)-mannanases, and β-(1,4)-xylanases involved in the breakdown of large polysaccharide chains (11–13). Whereas in these latter enzymes the substrate is polyhydroxylic and hydrophilic, in EGC it is amphiphilic: half hydrophilic, half hydrophobic. It is unclear what structural adaptations endow EGC with the ability to process such radically different substrates, particularly given their integral membrane localization. In the case of other glycolipid-hydrolyzing enzymes, such as those from lysosomes, activator proteins are known to play a role in "solubilizing" the substrates and presenting them to the active site. Indeed, the involvement of an activator protein has also been demonstrated for EGC (GenBank™ accession code E12780 (14)). *In vitro*, this activator can be replaced by detergents (14, 15). However, this is not feasible when using EGC to deglycosylate whole cells, as required in studies on the roles of endogenous gangliosides in biological systems. Consequently, commercial preparations of the enzyme are available to which the activator protein has been added.

The potential of gangliosides as therapies for cancer (16, 17), diabetes (18), and neurodegenerative diseases, such as Alzheimer (19, 20) and Parkinson diseases (21, 22), has prompted interest in the development of efficient, large scale syntheses to replace the current method of ganglioside isolation from contaminant-prone sources (*e.g.* bovine brain). Toward this end, protein engineers have recently harnessed EGC activity in

* This work was supported by Neose Technologies and by the Royal Society (United Kingdom), Howard Hughes Medical Institute, Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and the Protein Engineering Network of Centres of Excellence. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2OSW, 2OSX, and 2OSY) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada. Tel.: 604-822-0789; Fax: 604-822-5227; E-mail: natalie@byron.biochem.ubc.ca.

[2] The abbreviations used are: EGC, *endo*-glycoceramidase II; Lac, lactosyl/lactoside; G_{M2}AP, G_{M2} activator protein.

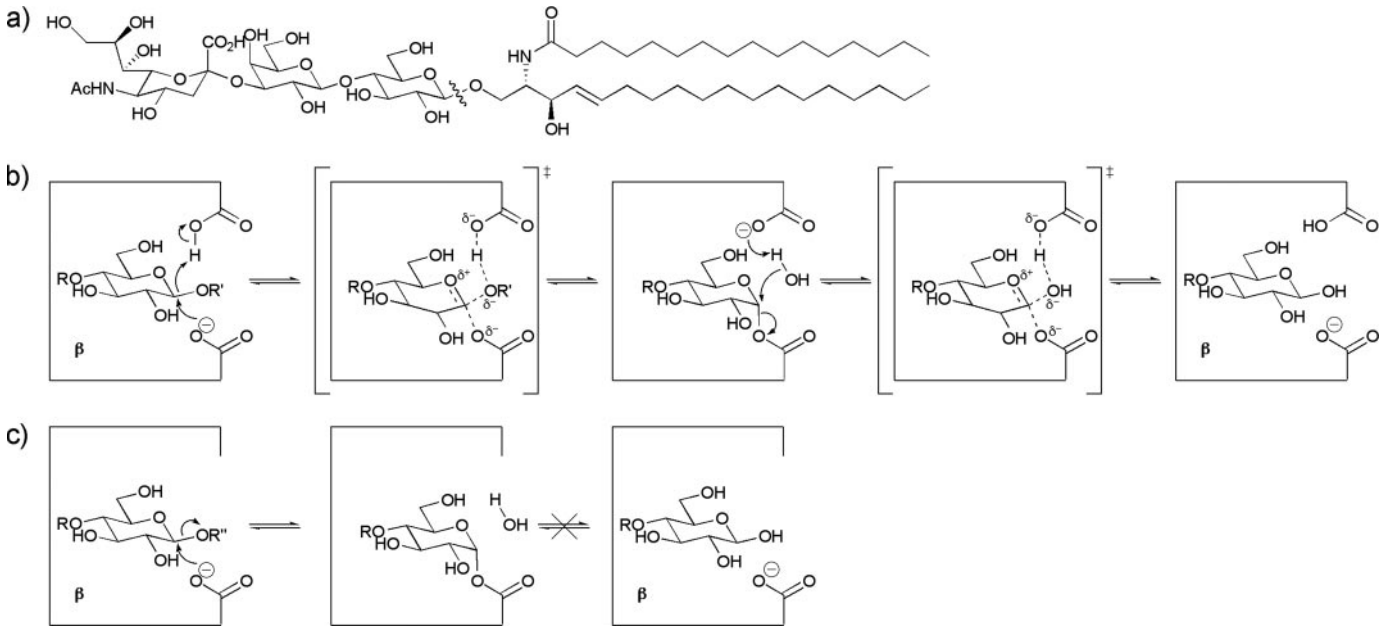

FIGURE 1. *a*, a typical ganglioside (G_{M3}) substrate for EGC. The position of EGC *endo*-hydrolysis is indicated by the *wavy line*. *b*, the hydrolysis mechanism of a retaining β-glycosidase (*R* and *R′* represent carbohydrate and ceramide moieties, respectively, in the case of EGC). *c*, the trapping of a glycosyl-enzyme intermediate with a general acid/base mutant (*R″* represents an activated leaving group).

order to address the difficulties in making these synthetically challenging molecules. By creating the glycosynthase (23, 24) of EGC, a nucleophile mutant of the enzyme capable of synthesizing but not hydrolyzing gangliosides, it was possible to synthesize a range of glycosphingolipids from their constituent oligosaccharide and lipid moieties in excellent yields (15). This latter study probed the ability of EGC to couple a broad range of oligosaccharide and lipid substrates, as did earlier complementary studies on the hydrolytic specificity of the wild-type enzyme (25). The structural basis of this specificity has yet to be ascertained.

Given the importance of gangliosides and their turnover in cellular regulation, it is surprising that the only structurally characterized glycolipid-processing enzyme to date is human glucocerebrosidase, whose deficiency results in Gaucher's disease (26, 27). This most likely arises from the inherent difficulties in handling membrane-associated proteins and hydrophobic substrates, such as ceramides.

Through trials of various detergents in crystallization and organic solvents to solubilize the substrates, we have succeeded in solving the x-ray crystal structures of apo-EGC, the Michaelis complex of the ganglioside substrate G_{M3} with a hydrolytically inactive nucleophile mutant, and a trapped glycosyl-enzyme intermediate. Together these structures detail the reaction coordinate, confirming the proposed double displacement mechanism and the identities of the catalytic glutamic acid residues. The G_{M3} complex, itself the product complex for the glycosynthase, details a novel glycosphingolipid binding topology, which provides a structural basis for the previously observed aglycone specificity of the enzyme. The crystal structures further offer valuable information for rational redesign of EGC toward the synthesis of novel ganglioside-derived therapeutics. This study represents a significant advance in our understanding of this unusual glycosidase activity, permitting insight into how glycosphingolipid hydrolysis is accomplished by this black sheep of the family 5 glycosidases.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The *Escherichia coli* codon-optimized gene for *endo*-glycoceramidase II from *Rhodococcus* sp. strain M-777, lacking the 30-residue N-terminal secretion signal sequence (*egc*), was synthesized (Blue Heron Biotechnology, Inc.) and subcloned into pET28a using the NdeI/XhoI restriction sites (15). *E. coli* BL21(Tuner) cells containing the *egc*/pET28a expression plasmid were grown at 37 °C to stationary phase in TYP medium containing 30 $\mu g \cdot ml^{-1}$ kanamycin. Protein production was induced by the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 0.1 mM following reduction of the temperature to 20 °C. The resulting His_6·EGC (residues 31–490) was purified by Ni(II) affinity chromatography to >95% purity as determined by SDS-PAGE analysis. For brevity, this recombinant protein is hereafter referred to as EGC.

The activity of EGC was confirmed by a colorimetric assay with 2,4-dinitrophenyl β-D-lactoside. Both E233A and E351S mutants of EGC were prepared as described previously (15) and purified using an identical protocol to that used for wild-type EGC. Selenomethionine-derivatized EGC was expressed using a metabolic inhibition protocol and M9 medium supplemented with 50 $mg \cdot dm^{-3}$ L-selenomethionine. Selenomethionine incorporation was observed >95% by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.

Crystallization—Crystals of apo-EGC were grown by the vapor diffusion method. Sitting drops of 1 μl of protein (10 $mg \cdot ml^{-1}$), containing 0.1% (v/v) Triton X-100, and 1 μl of well solution were equilibrated at 21 °C against a well solution of 20% (w/v) polyethylene glycol 3350, 0.175 M NaCl, and 0.1 M Tris-HCl, pH 8.5. Crystals belonging to the space group $P2_1$,

TABLE 1
Data collection and refinement statistics

	Selenomethionine	Native	EGC·G_{M3}	EGC·Lac
Data collection				
Space group	$P2_1$	$P2_1$	$C2$	$P2_1$
Cell dimensions				
a, b, c (Å)	53.4, 92.4, 94.6	53.8, 92.9, 94.5	77.8, 62.0, 102.8	53.9, 93.7, 94.5
α, β, γ (degrees)	90.0, 98.4, 90.0	90.0, 98.6, 90.0	90.0, 112.3, 90.0	90.0, 98.1, 90.0
Wavelength (Å)	0.980	1.542	1.000	1.542
Resolution range[a] (Å)	50.00-1.85 (1.92-1.85)	34.99-1.60 (1.69-1.60)	36.01-1.10 (1.16-1.10)	50.00-2.10 (2.18-2.10)
R_{sym}[a] (%)	8.7 (32.3)	5.3 (33.5)	5.4 (46.3)	6.1 (38.3)
$\langle I/\sigma(I)\rangle$[a]	17.8 (5.9)	22.4 (4.8)	14.7 (3.6)	21.5 (3.5)
No. of unique observations[a]	77,409 (7,663)	119,859 (17,253)	171,712 (23,927)	54,279 (5,402)
No. of total observations[a]	560,858	771,315 (111,092)	714,706 (99,776)	224,912
Completeness[a] (%)	99.9 (99.5)	99.2 (98.2)	93.8 (89.9)	99.7 (99.5)
Multiplicity[a]	7.2 (7.0)	6.4 (6.4)	4.2 (4.2)	4.1 (4.1)
Refinement				
R_{work}/R_{free} (%)		18.1/20.4	12.1/13.9	20.5/25.1
B-factors (Å^2)				
Protein		17.1	9.3	26.6
Ligand			9.3	24.8
Water		27.6	21.1	26.6
Root mean square deviations				
Bond lengths (Å)		0.006	0.010	0.014
Bond angles (degrees)		1.059	1.468	1.439

[a] Numbers in parentheses represent the value in the highest resolution shell.

with unit cell dimensions $a = 53.8$ Å, $b = 92.9$ Å, $c = 94.5$ Å, $\beta = 98.6°$, grew under these conditions over a period of 1 week. Co-crystallization with G_{M3} was accomplished by incubating EGC/E351S with G_{M3}, dissolved to a final concentration of 10 mM in a solution of 25% (w/v) polyethylene glycol 3350, 0.2 M NaCl, 1.0% (v/v) Triton X-100, 0.1 M Tris-HCl, pH 8.5, 10% (v/v) glycerol, and 10% (v/v) Me_2SO, in a 9:1 (v/v) ratio for 30 min at 21 °C before setting up sitting drop crystallization experiments as described above. Crystals belonging to the space group $C2$, with unit cell dimensions $a = 77.8$ Å, $b = 62.0$ Å, $c = 102.8$ Å, $\beta = 112.3°$ grew under these conditions over a period of 2 weeks. Cryoprotection was accomplished by soaking crystals in a solution of 25% (w/v) polyethylene glycol 3350, 0.2 M NaCl, 0.05% (v/v) Triton X-100, 0.1 M Tris-HCl, pH 8.5, and 10% (v/v) glycerol. Crystals were cryocooled by plunging into liquid nitrogen. Soaking experiments were carried out by incubating crystals of apo-EGC in solutions of cryoprotectant containing 2,4-dinitrophenyl β-D-lactoside at 10 mM for 10 min, before flash cooling by plunging into liquid nitrogen.

Data Collection and Structure Determination—X-ray data were collected at 100 K using a nitrogen stream. Data from selenomethionine-derivatized crystals and G_{M3} co-crystals were collected at Beamline 8.2.2 of the Advanced Light Source (Berkeley, CA), using an ADSC Q315 CCD detector. All other data were collected on an in-house rotating anode x-ray generator, using a Mar345 detector. The data were processed using either MOSFLM (28) and the CCP4 suite of programs (29), or DENZO and SCALEPACK (30).

18 selenium sites were located, and phases were calculated using the program SOLVE (31). The phases had a figure of merit of 0.40 over the resolution range to 2.0 Å, which increased to 0.72 following density modification with the program RESOLVE (32).

Refinement—An initial model incorporating ~65% of the structure was automatically built by RESOLVE (33) and manually completed using the program Coot (34). The model was subsequently transferred to the 1.6 Å data set for refinement. One cycle of simulated annealing was carried out using CNS (35). Further refinement, including individual B-factor refinement and the addition of water molecules, was carried out using the program REFMAC5 (36), with iterative manual rebuilding using Coot. 5% of the reflections were excluded for calculation of R_{free}. Non-crystallographic symmetry restraints were used throughout refinement. Two internal loops were omitted from the final model of the apo form of EGC due to poorly defined electron density. Additionally, the 12 N-terminal residues and upstream His_6 tag were not defined in the electron density and were omitted. Data collection and refinement statistics are shown in Table 1.

Structure Determination and Refinement of EGC Complexes—Initial phases for the $C2$ crystal form of EGC in complex with G_{M3} were calculated by molecular replacement using the program MOLREP (37). A monomer of EGC from the apo structure was used as the search model. The structure of the glycosyl-enzyme intermediate, resulting from the 2,4-dinitrophenyl β-D-lactoside soaking experiment, was solved using the native EGC structure as the initial source of phase information.

All structures of EGC complexes were refined using REFMAC5. Energy-minimized ligand models and their refinement restraints were created using the PRODRG server (38). Ligands were fitted to active site difference electron density, and further rounds of restrained refinement were carried out. Omit $mF_o - DF_c$ maps were calculated following removal of ligand atoms and random model perturbation using PDBSET.

All structures possessed one or more features in the electron density that, based upon coordination geometry and distance, were modeled as sodium ions originating from the crystallization conditions (39). Both the apo-EGC and the EGC·G_{M3} structures exhibited an unidentified chain of electron density at the periphery of the protein, distant from the active site that was unmodeled. This density resided on a 2-fold symmetry axis in the EGC·G_{M3} structure.

Final models were validated using MolProbity (40). Figures were prepared using PyMOL (41), electrostatic poten-

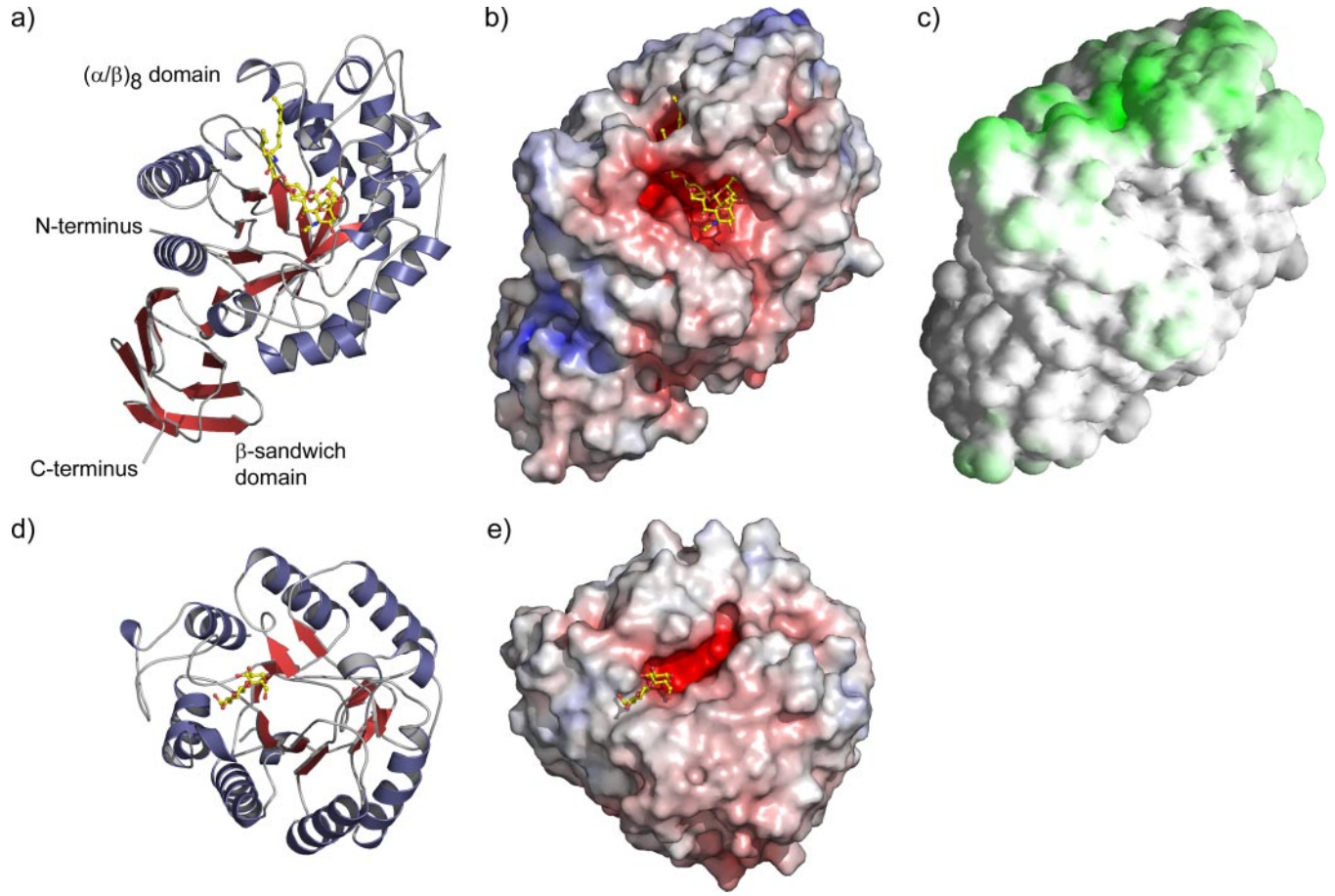

FIGURE 2. *a*, the structure of the EGC monomer. *b*, the electrostatic surface potential of EGC (*red*, electronegative; *blue*, electropositive; contoured from −15 to 1 *kT/e*). *c*, the hydrophobic surface potential of EGC (*green*, hydrophobic; *white*, polar). *d*, the structure of the β-(1,4)-glucanase from *Bacillus agaradherans*, Cel5A (Protein Data Bank code 2A3H). *e*, the electrostatic surface potential of Cel5A (*red*, electronegative; *blue*, electropositive; contoured from −20 to 1 *kT/e*). Bound ligands, G_{M3} (*a* and *b*) and cellobiose (*d* and *e*), are shown as *ball-and-stick representations* in *yellow*.

tials were calculated using APBS (42), and hydrophobic potentials were calculated using GRASP (43).

Enzyme Kinetics—Enzyme concentrations were measured by absorbance at 280 nm, using the extinction coefficient ϵ_{280} = 89,600 $cm^{-1} \cdot M^{-1}$, determined utilizing the methods of Gill and von Hipple (44). Michaelis-Menten kinetic parameters for the glycosynthase were determined by varying the concentration of either the donor or acceptor while utilizing a fixed concentration of 7 mM for the other species. Reactions were conducted at 37 °C in 25 mM sodium acetate, pH 5.0, containing 7% 1,2-dimethoxyethane. Reaction rates were measured using a VWR Symphony fluoride electrode interfaced with Logger Pro 2.2.1 analysis software (Vernier, Inc.). Before the addition of enzyme, the rate of spontaneous hydrolysis of the glycosyl fluoride was measured, and this rate was subsequently subtracted from the enzymatic rate. Glycosynthase reactions were initiated by the addition of EGC/E351S (100 μl) to a final concentration of 1–7 μM. Total reaction volume was 600 μl. Initial rates were plotted against substrate concentration, and Michaelis parameters were determined by non-linear regression analysis using Grafit 4.0 (Erithacus Software (45)). In cases where it was not possible to achieve enzyme saturation, k_{cat}/K_M values were determined from the gradient of the rate *versus* [S] plot.

RESULTS

Architecture of EGC—Recombinant EGC, lacking the 30-amino acid N-terminal signal sequence, was overexpressed in *E. coli* and purified by Ni(II) affinity chromatography. Diffraction quality EGC crystals were subsequently obtained in the presence of 0.05% (v/v) Triton X-100. Smaller EGC crystals were obtained with a number of other detergents, including octyl β-D-glucopyranoside, decyl β-D-maltopyranoside, and dodecyl β-D-maltopyranoside; however, no crystals were observed in the absence of detergent. The structure of EGC was determined by using the single-wavelength anomalous dispersion method with selenomethionine-derivatized protein. The EGC crystals belonged to the space group $P2_1$ and contained two molecules per asymmetric unit.

Each monomer of EGC is arranged in two distinct domains (Fig. 2*a*). The N-terminal domain (residues 43–407) assumes an $(\alpha/\beta)_8$ (triose-phosphate isomerase barrel) fold typical of the family 5 glycoside hydrolases (11–13). A subdomain (residues 137–195) forms a lid to the active site channel. The C-terminal domain (residues 408–490) assumes a β-sandwich fold, which resembles that of many carbohydrate-binding modules (46) and is composed of two sheets of four antiparallel β-strands. This domain is connected to the catalytic domain by a single

loop, although the relatively large buried surface area between the two (~900 Å^2) suggests a tight interaction rather than a flexible linkage. Its location on the opposite face of the $(\alpha/\beta)_8$ domain to the active site channel makes any role for this domain in binding of the carbohydrate portion of the substrate unlikely. Indeed, similar domains have been observed in the structures of a wide variety of glycosidases with no apparent involvement in substrate binding (47). These domains display no significant sequence similarities and hence are not detected by BLAST searches. Their presence in a number of other non-carbohydrate-related proteins adds further doubt to any specific role in carbohydrate association. It is possible this domain in EGC may simply stabilize the catalytic $(\alpha/\beta)_8$ domain.

Two molecules of EGC are present per asymmetric unit (root mean square deviation of 0.31 Å over 413 common C_α atoms between molecule A and molecule B); however, the small interaction area between the two molecules (a buried surface of ~350 Å^2) and static light scattering analysis (data not shown) suggest that the observed dimer results from crystallographic packing only, with the monomeric form being representative of the physiological state of EGC.

EGC Active Site Structure—The EGC active site channel is formed by the loops of the $(\alpha/\beta)_8$ domain and consists of two distinct regions. To one side of the predicted catalytic residues, the active site channel is broad (~14 Å) and lined mainly with polar residues (Fig. 2*b*). A highly disordered loop (residues 145–154) is positioned to one side of this channel. The active site narrows on the opposite side of the catalytic site to an ~8 Å channel lined predominantly with hydrophobic residues. This channel subsequently opens onto a distinctly flat surface of the enzyme, which also appears largely composed of hydrophobic residues (Fig. 2, *b* and *c*). A further disordered loop (residues 311–315) is positioned to the side of the hydrophobic channel. The side chains of the catalytic residues, Glu^{233} and Glu^{351} (predicted by sequence homology to other family 5 glycosidases), are separated by an average distance of ~5.2 Å, consistent with their role in a double displacement mechanism with net retention of stereochemistry (48). The side chains are oriented in an approximately perpendicular manner, which, when inspected in the presence of bound G_{M3} (see below), conforms to the *anti*-protonation trajectory of the general acid/base, again common to other family 5 glycosidases.

Inspection of the mF_o − DF_c map of the active site revealed density attributable to a molecule of Tris bound between the predicted catalytic residues. The Tris molecule is coordinated by the side chains of Lys^{66}, Asp^{137}, Asn^{232}, Glu^{233}, and Glu^{351} and appears to partially mimic the coordination of a sugar bound in the −1 subsite.

EGC·G_{M3} Substrate Complex—In order to prevent substrate hydrolysis, crystallization experiments with the substrate G_{M3} were carried out with a nucleophile knock-out mutant, E351S. This is the same mutant as that which was employed as the glycosynthase in previous studies. Soaking EGC/E351S crystals with G_{M3} led to a significant deterioration in diffraction quality. EGC/E351S was therefore co-crystallized with G_{M3}. EGC·G_{M3} crystals belonged to the space group *C*2 and contained one molecule per asymmetric unit. The EGC·G_{M3} structure was solved by molecular replacement to a resolution of 1.1 Å. Inspection of the mF_o − DF_c map revealed density attributable to G_{M3} bound in the proposed active site channel. G_{M3} was modeled and refined to unitary occupancy, with an average *B*-factor of 9.3 Å^2. There was no indication of residual Tris binding.

The overall structure and fold of EGC appears broadly unchanged following binding of G_{M3}, as reflected in a root mean square deviation of 0.47 Å over 404 common C_α atoms between the apo and G_{M3}-bound forms. Both unmodeled loops of the apo-EGC structure are, however, well defined in the electron density of the EGC·G_{M3} structure. Notably, the loop 145–151 appears to be dynamic and has been modeled by two alternate main-chain conformations.

The three pyranoside rings of G_{M3} fold into a distinctive curve (Fig. 3), with the 6-OH of glucose interacting directly with the 9-OH of sialic acid over a distance of 2.8 Å. Notably, this forms the only non-solvent-binding interaction for the sialic acid moiety. The coordination of the G_{M3} is described in detail in Fig. 3*c*. In common with many carbohydrate-binding proteins, the substrate binding site of EGC is lined with a number of aromatic amino acids that form hydrophobic interactions with the sugar rings; the glucose of G_{M3} interacts with the side chains of Tyr^{306} and Trp^{382}, whereas the galactose ring interacts with the side chain of Trp^{178}. The side chain of Trp^{389} constitutes an additional boundary to the active site cavity. There is no observable distortion of the G_{M3} glucose moiety from a chair conformation, as might have been expected in a Michaelis complex (49), suggesting that the side chain of the nucleophile Glu^{351} may be required for substrate distortion.

Both hydrocarbon chains of the G_{M3} ceramide are well defined in the electron density. The electron density deteriorates after the chains exit the enzyme active site channel, presumably due to increased flexibility. The acyl and sphingosine chains of the ceramide are stacked vertically with respect to the hydrophobic channel, with the acyl chain located below the sphingosine chain. The ceramide-binding channel is lined by the hydrophobic side chains of amino acids Leu^{180}, Tyr^{182}, Ile^{183}, Phe^{235}, Ile^{276}, and Leu^{308}. Upon binding of G_{M3}, the side chains of Arg^{177} and Asp^{311} localize to form a "cap" over the ceramide-binding channel. Presumably as a consequence of this conformational change, the flexible 311–315 loop region becomes well defined in the electron density. Furthermore, there is a slight conformational change in part of the lid subdomain (residues 166–193) to accommodate the change in Arg^{177}.

EGC·Lac Glycosyl-Enzyme Intermediate—The glycosyl-enzyme intermediate was trapped by soaking crystals of the general acid/base knock-out mutant, EGC/E233A, with an activated sugar donor, 2,4-dinitrophenyl β-D-lactoside. The presence of such a good leaving group overcomes the lack of general acid catalysis, allowing formation of the glycosyl-enzyme intermediate. Conversely, the absence of the catalytic base residue slows hydrolysis sufficiently to allow accumulation of the intermediate (Fig. 1*c*). Cryogenic protection then preserves this state during data collection. Inspection of the active sites of the resulting structure revealed electron density consistent with the presence of a lactosyl moiety covalently bound to the side chain of the catalytic nucleophile, Glu^{351} (Fig. 4).

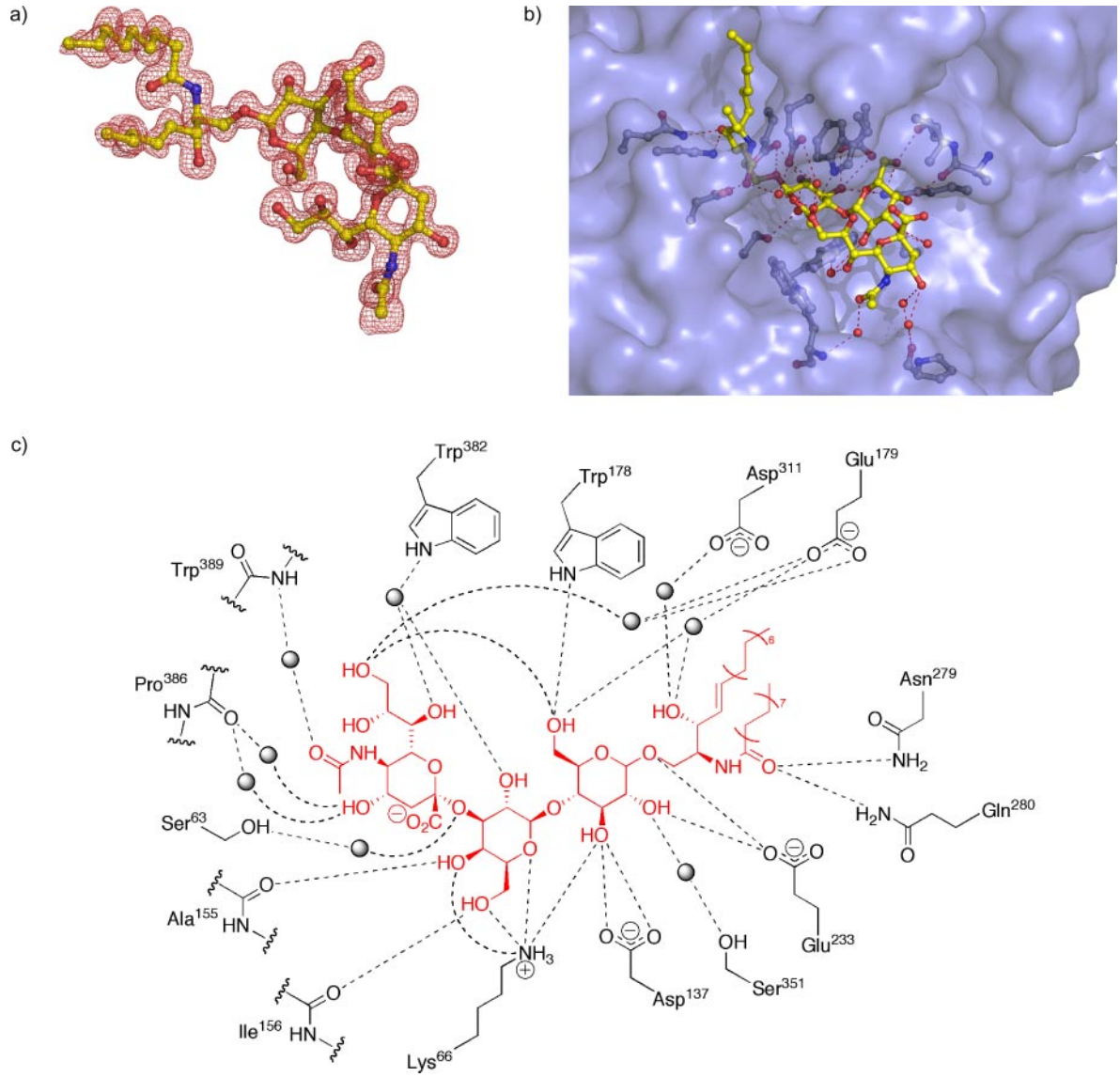

FIGURE 3. *a*, electron density for the bound G_{M3}. An mF_o − DF_c (36) electron density map, calculated after random model perturbation and refinement with G_{M3} atoms omitted, is shown contoured around the G_{M3} at 2.5 σ in *red*. *b*, a surface representation depicting the G_{M3} binding site. G_{M3} is shown as a *ball-and-stick representation* in *yellow*, surrounded by its ligands in the active site of EGC. *c*, a schematic representation of polar, close contacts involved in the binding of G_{M3}. Water molecules are represented by *gray spheres*.

There was no evidence for the binding of the 2,4-dinitrophenolate product. The modeled glycosyl-enzyme intermediate was refined to unitary occupancy, with an average *B*-factor of 24.8 Å^2. The glucosyl and galactosyl moieties of the intermediate are bound in conformations comparable with those of G_{M3} and possess similar interactions with the enzyme. There are no discernable conformational differences between the EGC·Lac and the apo-EGC structure. As with the apo structure, however, the ceramide-binding regions, in particular the side chain of Arg^{177} and the 308–317 loop, are poorly defined in the EGC·Lac structure and have not been modeled. The identification of the glycosyl-enzyme intermediate confirms the identity of Glu^{351} and Glu^{233} as the catalytic nucleophile and general acid/base, respectively.

Attempts to soak EGC/E233A crystals with 2,4-dinitrophenyl β-D-sialyllactoside gave rise to electron density suggestive of a sialyllactosyl glycosyl-enzyme intermediate, but only in low occupancy. Presumably, crystal contacts restrict the soaking of this larger trisaccharide, and the poor occupancy prevented confident modeling of this complex.

Probing the Ceramide-binding Site—To further assess the basis of lipid specificity in EGC, the kinetic parameters of various lipid acceptors for the synthetic glycosynthase reaction were determined. These are presented in Table 2.

DISCUSSION

In common with the structures of other family 5 glycosidases, the catalytic domain of EGC forms an $(\alpha/\beta)_8$ fold, with the catalytic center positioned in a channel on the upper surface of the barrel. It is the constitution of the EGC active site channel, however, which distinguishes it markedly from other members of family 5 and provides insight into the unusual substrate

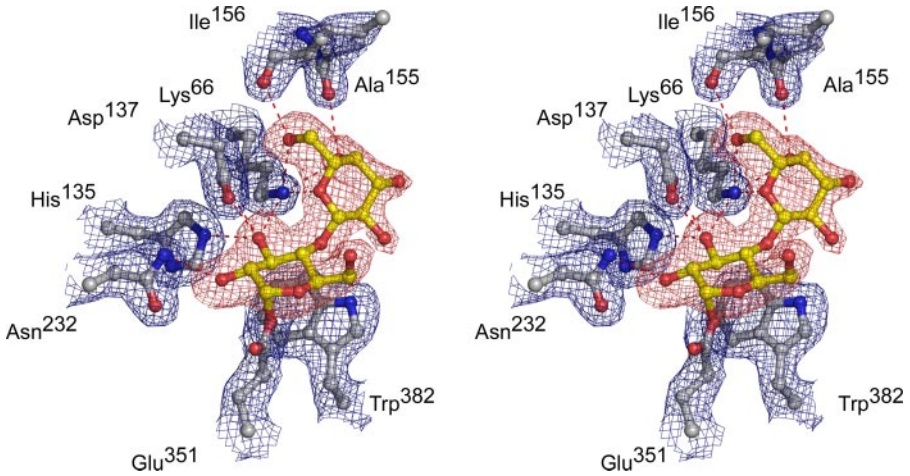

FIGURE 4. **A stereo representation of the glycosyl-enzyme intermediate derived from 2,4-dinitrophenyl β-D-lactoside soaking experiments.** The covalently bound lactosyl moiety is shown as a *ball-and-stick representation* in *yellow*, surrounded by its ligands in the active site of EGC. An m$F_o - DF_c$ (36) electron density map, calculated after random model perturbation and refinement with lactosyl atoms omitted, is shown contoured around the lactosyl moiety at 2.5 σ in *red*. A 2m$F_o - DF_c$ electron density map is shown contoured around the protein ligands at 1.0 σ in *blue*.

TABLE 2
Kinetic parameters for the glycosynthase reaction of EGC/E351S

Acceptor	k_{cat}	K_M	k_{cat}/K_M
	s^{-1}	mM	$M^{-1}\cdot s^{-1}$
D-*erythro*-Sphingosine	0.07 ± 0.04	0.04 ± 0.01	1900 ± 200
D-*erythro*-Dihydrosphingosine	0.12 ± 0.05	0.3 ± 0.1	340 ± 20
L-*threo*-Sphingosine			1.3 ± 0.3
L-*threo*-Dihydrosphingosine			1.1 ± 0.2
Phytosphingosine			0.17 ± 0.05

selectivity exhibited by this enzyme. Unlike those family 5 glycosidases that have adapted to accept entirely polar polysaccharides, the active site of EGC displays a split personality to accommodate both the hydrophilic sugar and hydrophobic ceramide components of its glycosphingolipid substrates.

The glycone binding site described by the EGC·G_{M3} complex, on the nonreducing (minus subsites) side of the catalytic center, exhibits many of the features common to sugar-binding proteins. These include many polar contacts between enzyme and ligand to "solvate" the hydroxyls and a number of hydrophobic interactions with aromatic side chains that further stabilize the bound sugars. This cavity is broader, however, than the linear channels typically observed in the β-glucanases, β-mannanases, and β-xylanases of family 5 (see comparison of EGC and Cel5A β-glucanase (11) (Fig. 2)), presumably to allow binding of the branched sugar moieties that EGC is known to process. The flexible loop region (residues 145–156) to the side of the active site is positioned close to the axial 4-OH of the G_{M3} galactose moiety, the location where the additional sugar groups branch off in the more extended G_{M1} and G_{M2} gangliosides. Indeed, studies probing substrate glycone selectivity suggested only slightly lower tolerance of the larger G_{M1} and G_{M2} substrates (25), and these larger oligosaccharides were seen to be efficiently processed by the glycosynthase mutant (15). It is possible, therefore, that this loop may facilitate an induced fit mechanism, allowing expansion of the active site for these alternative substrates. The striking curve defined by the three sugar moieties of G_{M3} may be a determinant of the glycone selectivity of EGC; it is possible that an intramolecular self-association of the substrate into a predefined shape assists recognition and binding.

To accommodate the binding of ceramide-based aglycone moieties, the active site channel downstream of the catalytic center is lined with hydrophobic residues. Interestingly, the lipid-binding topology observed in the EGC·G_{M3} complex differs from that of other glycosphingolipid-binding proteins, which typically encapsulate the hydrophobic tails, revealing only the polar sugars and lipid head groups to solvent. For the CD1 family of antigen-presenting proteins (50–59), although each member exhibits specific adaptations consistent with varying substrate specificities, the overall protein architecture and mode of lipid binding is similar; the lipid tails are fully enclosed within complex networks of hydrophobic channels specific to the length and shape of the lipid chains. The polar sugars and lipid head groups are hydrogen-bonded at the protein surface for presentation to T-cell receptors. Similarly, the structures of human and bovine glycolipid transfer proteins (60, 61) reveal analogous modes of substrate binding; the polar glycone moieties are positioned at the protein surface, whereas the lipid tails are bound within a single hydrophobic tunnel completely enclosed within the protein environment. Structural studies of the membrane-lipid activator proteins saposin B (62) and human G_{M2} activator protein (G_{M2}AP) (63, 64) further exemplify lipid substrate binding by encapsulation within hydrophobic protein cores. All these classes of glycosphingolipid-binding protein, regardless of function, thus appear to utilize a comparable glycosphingolipid binding pocket. In contrast, the structure of the EGC·G_{M3} complex reveals a novel lipid binding topology, which most probably reflects the requirement for EGC to more intimately associate with both polar and apolar components of its substrates to facilitate *endo*-hydrolysis. It is also possible that this topology allows the association of the activator protein employed by this enzyme.

The EGC activator protein is considerably larger (~70 kDa) than the mammalian activators, human G_{M2}AP (~20 kDa) and the saposin family (~9 kDa). In the absence of structural information on the EGC activator, however, the mechanism of lipid presentation to the enzyme is unclear. Analogous to hypotheses for G_{M2}AP and saposin B membrane-lipid activation, the lipid may be extracted from the membrane to permit enzyme interaction, or the membrane may be modified to enable lipid-enzyme association. It is possible that the flat hydrophobic surface of EGC (Fig. 2, *b* and *c*) observed near the ceramide exit tunnel may facilitate either direct interaction with the membrane surface or the formation of a protein-protein complex with the activator.

An interesting model for the interaction of an activator pro-

tein with its cognate enzyme is that deduced for the human β-hexosaminidase A. This is a family 20 glycosidase that catalyzes G_{M2}AP-dependent *exo*-hydrolysis of the terminal *N*-acetyl-D-galactosamine of G_{M2} to yield G_{M3}. Predictive models of a β-hexosaminidase A·G_{M2}AP·G_{M2} complex based on mutagenesis analysis (65) suggest binding of only the sugar moiety by β-hexosaminidase A, whereas the hydrophobic lipid tails remain bound by G_{M2}AP. Hence, in comparison with EGC, a lipid binding site is not required by β-hexosaminidase A, reflecting the significant differences in substrate binding imposed by *exo*- or *endo*-hydrolysis.

The unique lipid-binding topology exhibited by EGC is manifested in the high degree of selectivity for lipid substrates with the correct head group structure. In the glycosynthase-catalyzed glycosylation of sphingolipids, a dramatic reduction in catalytic efficiency is observed when modified lipids are employed. The measured K_M value for D-*erythro*-sphingosine was 40 μM, which was increased by 10-fold for the hydrogenated analog D-*erythro*-dihydrosphingosine, although k_{cat} was roughly the same for both compounds, 0.1 s^{-1}. When more substantial changes were introduced into the head group, the catalytic efficiency was dramatically reduced, making enzyme saturation impossible. When L-*threo*-sphingosine and its hydrogenated analog, L-*threo*-dihydrosphingosine, in which both stereocenters of the head group are inverted relative to D-*erythro*-sphingosine, were employed, a 2000-fold reduction in k_{cat}/K_M was observed. These observations illustrate the strict stereoselectivity of EGC toward the lipid head group. These restraints are consistent with the observed specific bond interactions and steric constraints imposed by the substrate-binding channel on the ceramide. Hydration of the double bond in D-*erythro*-sphingosine, giving phytosphingosine, proved to be even more catastrophic, resulting in a 10,000-fold reduction in k_{cat}/K_M.

Combined, these structural and biochemical studies illustrate the essential adaptations of EGC from the typical family 5 scaffold. The broad glycan tolerance is explained by the wide and potentially flexible sugar-binding site; the narrow hydrophobic ceramide-binding channel enables intimate association with the ganglioside and thus *endo*-hydrolysis; and encapsulation of the lipid ensures close contacts with the lipid head group facilitating stereoselectivity. Furthermore, the structure permits speculation on the potential membrane and activator protein interactions that allow the enzyme to function on membrane-associated substrates. Importantly, the structural knowledge also opens the possibilities of rational design of EGC glycosynthase substrate specificity toward the synthesis of non-natural, therapeutically useful glycosphingolipids.

Acknowledgments—We thank Dr. Andrew Lovering and Dr. Paula Lario for assistance with data collection, Dr. Calvin Yip for static light scattering analysis, Dr. Bernard Henrissat for helpful discussion, and the staff at Beamline 8.2.2 of the Advanced Light Source (Berkeley, CA) for expert assistance. We also thank Neose Technologies for providing reagents.

REFERENCES

1. Tettamanti, G. (2004) *Glycoconj. J.* **20,** 301–317
2. Hakomori, S. (1981) *Annu. Rev. Biochem.* **50,** 733–764
3. Ashida, H., Yamamoto, K., Kumagai, H., and Tochikura, T. (1992) *Eur. J. Biochem.* **205,** 729–735
4. Basu, S. S., Dastgheib-Hosseini, S., Hoover, G., Li, Z., and Basu, S. (1994) *Anal. Biochem.* **222,** 270–274
5. Horibata, Y., Okino, N., Ichinose, S., Omori, A., and Ito, M. (2000) *J. Biol. Chem.* **275,** 31297–31304
6. Horibata, Y., Sakaguchi, K., Okino, N., Iida, H., Inagaki, M., Fujisawa, T., Hama, Y., and Ito, M. (2004) *J. Biol. Chem.* **279,** 33379–33389
7. Ito, M., and Yamagata, T. (1986) *J. Biol. Chem.* **261,** 14278–14282
8. Li, Y.-T., Ishikawa, Y., and Li, S.-C. (1987) *Biochem. Biophys. Res. Commun.* **149,** 167–172
9. Zhou, B., Li, S.-C., Laine, R. A., Huang, R. T. C., and Li, Y.-T. (1989) *J. Biol. Chem.* **264,** 12272–12277
10. Zechel, D. L., and Withers, S. G. (2000) *Acc. Chem. Res.* **33,** 11–18
11. Davies, G. J., Dauter, M., Brzozowski, A. M., Bjornvad, M. E., Andersen, K. V., and Schulein, M. (1998) *Biochemistry* **37,** 1926–1932
12. Hilge, M., Gloor, S. M., Rypniewski, W., Sauer, O., Heightman, T. D., Zimmermann, W., Winterhalter, K., and Piontek, K. (1998) *Structure* **6,** 1433–1444
13. Larson, S. B., Day, J., Barba de la Rosa, A. P., Keen, N. T., and McPherson, A. (2003) *Biochemistry* **42,** 8411–8422
14. Ito, M., Ikegami, Y., and Yamagata, T. (1991) *J. Biol. Chem.* **266,** 7919–7926
15. Vaughan, M. D., Johnson, K., DeFrees, S., Tang, X., Warren, R. A. J., and Withers, S. G. (2006) *J. Am. Chem. Soc.* **128,** 6300–6301
16. Hakomori, S. (2000) *Glycoconj. J.* **17,** 627–647
17. Nojiri, H., Yamana, H., Shirouzu, G., Suzuki, T., and Isono, H. (2002) *Cancer Detect. Prev.* **26,** 114–120
18. Allende, M. L., and Proia, R. L. (2002) *Curr. Opin. Struct. Biol.* **12,** 587–592
19. Matsuoka, Y., Saito, M., LaFrancois, J., Saito, M., Gaynor, K., Olm, V., Wang, L., Casey, E., Lu, Y., Shiratori, C., Lemere, C., and Duff, K. (2003) *J. Neurosci.* **23,** 29–33
20. Svennerholm, L., Brane, G., Karlsson, I., Lekman, A., Ramstrom, I., and Wikkelso, C. (2002) *Dement. Geriatr. Cogn.* **14,** 128–136
21. Pope-Coleman, A., and Schneider, J. S. (1998) *Rest. Neurol. Neurosci.* **12,** 255–266
22. Itoh, M., Fukumoto, S., Iwamoto, T., Mizuno, A., Rokutanda, A., Ishida, H. K., Kiso, M., and Furukawa, K. (2001) *Glycobiology* **11,** 125–130
23. Hancock, S. M., Vaughan, M. D., and Withers, S. G. (2006) *Curr. Opin. Chem. Biol.* **10,** 509–519
24. Mackenzie, L. F., Wang, Q. P., Warren, R. A. J., and Withers, S. G. (1998) *J. Am. Chem. Soc.* **120,** 5583–5584
25. Izu, H., Izumi, Y., Kurome, Y., Sano, M., Kondo, A., Kato, I., and Ito, M. (1997) *J. Biol. Chem.* **272,** 19846–19850
26. Dvir, H., Harel, M., McCarthy, A. A., Toker, L., Silman, I., Futerman, A. H., and Sussman, J. L. (2003) *EMBO Rep.* **4,** 704–709
27. Premkumar, L., Sawkar, A. R., Boldin-Adamsky, S., Toker, L., Silman, I., Kelly, J. W., Futerman, A. H., and Sussman, J. L. (2005) *J. Biol. Chem.* **280,** 23815–23819
28. Leslie, A. G. W. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 1696–1702
29. Collaborative Computational Project 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
30. Otwinowski, Z., and Minor, W. (1997) *Method. Enzymol.* **276,** 307–326
31. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 849–861
32. Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 965–972
33. Terwilliger, T. C. (2003) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59,** 38–44
34. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132
35. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S.,

Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
36. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 247–255
37. Vagin, A. A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30,** 1022–1025
38. Schuttelkopf, A. W., and van Aalten, D. M. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 1355–1363
39. Harding, M. M. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 872–874
40. Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins Struct. Funct. Bioinf.* **50,** 437–450
41. DeLano, W. L. (2002) *The PyMol Molecular Graphics System*, Version 0.99, DeLano Scientific, San Carlos, CA
42. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10037–10041
43. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11,** 281–296
44. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem.* **182,** 319–326
45. Leatherbarrow, R. J. (2004) *Grafit*, Version 4.0.21, Erithacus Software Ltd., Horley, UK
46. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) *Biochem. J.* **382,** 769–781
47. Sulzenbacher, G., Bignon, C., Nishimura, T., Tarling, C. A., Withers, S. G., Henrissat, B., and Bourne, Y. (2004) *J. Biol. Chem.* **279,** 13119–13128
48. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* **4,** 885–892
49. Vasella, A., Davies, G. J., and Bohm, M. (2002) *Curr. Opin. Chem. Biol.* **6,** 619–629
50. Batuwangala, T., Shepherd, D., Gadola, S. D., Gibson, K. J., Zaccai, N. R., Fersht, A. R., Besra, G. S., Cerundolo, V., and Jones, E. Y. (2004) *J. Immunol.* **172,** 2382–2388
51. Gadola, S. D., Zaccai, N. R., Harlos, K., Shepherd, D., Castro-Palomino, J. C., Ritter, G., Schmidt, R. R., Jones, E. Y., and Cerundolo, V. (2002) *Nat. Immunol.* **3,** 721–726
52. Giabbai, B., Sidobre, S., Crispin, M. D., Sanchez-Ruiz, Y., Bachi, A., Kronenberg, M., Wilson, I. A., and Degano, M. (2005) *J. Immunol.* **175,** 977–984
53. Koch, M., Stronge, V. S., Shepherd, D., Gadola, S. D., Mathew, B., Ritter, G., Fersht, A. R., Besra, G. S., Schmidt, R. R., Jones, E. Y., and Cerundolo, V. (2005) *Nat. Immunol.* **6,** 819–826
54. Wu, D., Zajonc, D. M., Fujio, M., Sullivan, B. A., Kinjo, Y., Kronenberg, M., Wilson, I. A., and Wong, C. H. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 3972–3977
55. Zajonc, D. M., Cantu, C., III, Mattner, J., Zhou, D., Savage, P. B., Bendelac, A., Wilson, I. A., and Teyton, L. (2005) *Nat. Immunol.* **6,** 810–818
56. Zajonc, D. M., Crispin, M. D., Bowden, T. A., Young, D. C., Cheng, T. Y., Hu, J., Costello, C. E., Rudd, P. M., Dwek, R. A., Miller, M. J., Brenner, M. B., Moody, D. B., and Wilson, I. A. (2005) *Immunity* **22,** 209–219
57. Zajonc, D. M., Elsliger, M. A., Teyton, L., and Wilson, I. A. (2003) *Nat. Immunol.* **4,** 808–815
58. Zajonc, D. M., Maricic, I., Wu, D., Halder, R., Roy, K., Wong, C. H., Kumar, V., and Wilson, I. A. (2005) *J. Exp. Med.* **202,** 1517–1526
59. Zeng, Z., Castano, A. R., Segelke, B. W., Stura, E. A., Peterson, P. A., and Wilson, I. A. (1997) *Science* **277,** 339–345
60. Airenne, T. T., Kidron, H., Nymalm, Y., Nylund, M., West, G., Mattjus, P., and Salminen, T. A. (2006) *J. Mol. Biol.* **355,** 224–236
61. Malinina, L., Malakhova, M. L., Teplov, A., Brown, R. E., and Patel, D. J. (2004) *Nature* **430,** 1048–1053
62. Ahn, V. E., Faull, K. F., Whitelegge, J. P., Fluharty, A. L., and Prive, G. G. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 38–43
63. Wright, C. S., Li, S. C., and Rastinejad, F. (2000) *J. Mol. Biol.* **304,** 411–422
64. Wright, C. S., Zhao, Q., and Rastinejad, F. (2003) *J. Mol. Biol.* **331,** 951–964
65. Lemieux, M. J., Mark, B. L., Cherney, M. M., Withers, S. G., Mahuran, D. J., and James, M. N. (2006) *J. Mol. Biol.* **359,** 913–929