Structural and Mechanistic Analyses of endo-Glycoceramidase II, a Membrane-associated Family 5 Glycosidase in the Apo and G\textsubscript{M\textsubscript{3}} 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.M61455200

Matthew E. C. Caines \textsuperscript{1}, Mark D. Vaughan \textsuperscript{1}, Chris A. Tarling \textsuperscript{3}, Susan M. Hancock \textsuperscript{1}, R. Antony J. Warren \textsuperscript{6}, Stephen G. Withers \textsuperscript{4}\textsuperscript{5}, and Natalie C. J. Strynadka \textsuperscript{11}

From the Departments of \textsuperscript{1}Biochemistry and Molecular Biology, \textsuperscript{6}Chemistry, and \textsuperscript{11}Microbiology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

\textsuperscript{1} 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.

end\textsubscript{o}-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 \textit{Rhodococcus} sp., in the apo form, in complex with the ganglioside \textit{G}\textsubscript{M\textsubscript{3}} (Svennerholm ganglioside nomenclature (Svennerholm, L. (1964) \textit{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. 1a) 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)\textsuperscript{2} from \textit{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. 1b) (10).

EGC is noticeably distinct in function from other family 5 glycosidases, including the prototypical \(\beta\)-(1,4)-glucanases (cellulases), \(\beta\)-(1,4)-mannanases, and \(\beta\)-(1,4)-xylanases involved in the breakdown of large polysaccharide chains (11–13). Whereas in these latter enzymes the substrate is polyhydroxyl 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\textsuperscript{TM} accession code E12780 (14)). \textit{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

\footnotesize{\textsuperscript{2} The abbreviations used are: EGC, endo-glycoceramidase II; Lac, lactosyl/lactoside; \textit{G}M\textsubscript{4}AP, \textit{G}G\textsubscript{4} activator protein.}
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 GM3 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 GM3 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 μg/ml 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 His6 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 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·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 P2₁.
with unit cell dimensions $a = 53.8 \text{ Å}, b = 92.9 \text{ Å}, c = 94.5 \text{ Å}$, $\beta = 98.6^\circ$, grew under these conditions over a period of 1 week. Co-crystallization with $G_{M3}$ was accomplished by incubating EGC/E3515 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$_2$SO, in a 9:1 (v/v) ratio for 30 min at 21 °C. The crystals were subsequently transferred to the 1.6 Å data set for refinement.

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 at Beamline 8.2.2 of the Advanced Light Source (Berkeley, CA), using an ADSC Q315 CCD detector. All other 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 at Beamline 8.2.2 of the Advanced Light Source (Berkeley, CA), using an ADSC Q315 CCD detector.

Table 1: Data collection and refinement statistics

| Data collection       | Selenomethionine | Native | EGC-G$_{M3}$ | EGC-Lac |
|-----------------------|------------------|--------|--------------|---------|
| Space group           | $P2_1$           | $P2_1$ |             |         |
| 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 |
| $a, b, c$ (degrees)   | 90.0, 98.4, 90.0 | 90.0, 98.6, 90.0 | 90.0, 112.3, 90.0 | 90.8, 91.0, 90.0 |
| Resolution range (Å) | 0.980            | 1.542  | 1.000        | 1.542   |
| Resolution range (Å²)| 34.99-1.60       | 36.01-1.10 | 50.00-2.10  |         |
| $R_{merge}$ (%)       | 8.7 (32.3)       | 5.3 (33.5) | 5.4 (46.3)  |         |
| $(I/\sigma(I))$       | 17.8 (5.9)       | 22.4 (4.8) | 14.7 (3.6)  |         |
| No. of unique observations | 77,409 (7,663) | 119,859 (12,753) | 171,712 (23,927) | 54,279 (5,402) |
| No. of total observations | 560,858       | 771,315 (111,092) | 714,706 (99,776) | 224,912 |
| Completeness (%)      | 99.9 (99.5)      | 99.2 (98.2) | 93.8 (89.9) | 99.7 (99.5) |
| Multiplicity (%)      | 7.2 (7.0)        | 6.4 (6.4)  | 4.2 (4.2)  | 4.1 (4.1) |

$^a$ Numbers in parentheses represent the value in the highest resolution shell.

Refinement

| Refinement | Selenomethionine | Native | EGC-G$_{M3}$ | EGC-Lac |
|------------|------------------|--------|--------------|---------|
| $R_{factor}$ | 18.1/20.4       | 12.1/13.9 | 20.5/25.1   |         |
| $B$-factors (Å²) | 17.1       | 9.3     | 26.6         |         |
| Ligand     | 9.3             |         | 24.8         |         |
| Water      | 27.6            | 21.1    | 26.6         |         |
| Root mean square deviations | 0.006       | 0.010   | 0.014        |         |
| Bond lengths (Å) | 1.059       | 1.468   | 1.439        |         |

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 $R_{merge}$. 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-$\beta$-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 m$F_o$ – DF$e$ maps were calculated following removal of ligand atoms and random model perturbation using PDBeSET.

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-
Structural Analyses of endo-Glycoceramidase II

Enzyme Kinetics—Enzyme concentrations were measured by absorbance at 280 nm, using the extinction coefficient $\varepsilon_{280} = 89,600 \text{ cm}^{-1} \cdot \text{M}^{-1}$, determined utilizing the methods of Gill and von Hippel (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-di-methoxyethane. 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 μM) 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 $P_{2_1}2_1$ and contained two molecules per asymmetric unit.

Each monomer of EGC is arranged in two distinct domains (Fig. 2a). The N-terminal domain (residues 43–407) assumes an (α/β)$_8$ (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

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, GM3 (a and b) and cellobiose (d and e), are shown as ball-and-stick representations in yellow.
Structural Analyses of endo-Glycoceramidase II

loop, although the relatively large buried surface area between the two (∼900 Å²) suggests a tight interaction rather than a flexible linkage. Its location on the opposite face of the (α/β)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 (α/β)₈ domain.

Two molecules of EGC are present per asymmetric unit (root mean square deviation of 0.31 Å over 413 common Cᵦ atoms between molecule A and molecule B); however, the small interaction area between the two molecules (a buried surface of ∼350 Å²) 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 (α/β)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. 2b). 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 on the side of the hydrophobic channel. The side chains of the catalytic residues, Glu٢٣٣ and Glu٣٥١ (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 G₃M₃ (see below), conforms to the anti-protonation trajectory of the general acid/base, again common to other family 5 glycosidases.

Inspection of the mF₀ – DF₀ 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 Lys٦٦, Asp١٣٧, Asn٢٣٢, Glu٢٣٣, and Glu٣٥١ and appears to partially mimic the coordination of a sugar bound in the −1 subsite.

EGC-G₃M₃ Substrate Complex—In order to prevent substrate hydrolysis, crystallization experiments with the substrate G₃M₃ 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 G₃M₃ led to a significant deterioration in diffraction quality. EGC/E351S was therefore co-crystallized with G₃M₃. EGC-G₃M₃ crystals belonged to the space group C₂ and contained one molecule per asymmetric unit. The EGC-G₃M₃ structure was solved by molecular replacement to a resolution of 1.1 Å. Inspection of the mF₀ – DF₀ map revealed density attributable to G₃M₃ bound in the proposed active site channel. G₃M₃ was modeled and refined to unitary occupancy, with an average B-factor of 9.3 Å². There was no indication of residual Tris binding.

The overall structure and fold of EGC appears broadly unchanged following binding of G₃M₃, as reflected in a root mean square deviation of 0.47 Å over 404 common Cᵦ atoms between the apo and G₃M₃-bound forms. Both unmodeled loops of the apo-EGC structure are, however, well defined in the electron density of the EGC-G₃M₃ 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 G₃M₃ 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 G₃M₃ is described in detail in Fig. 3c. 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 G₃M₃ interacts with the side chains of Tyr٣٦٦ and Trp٣٨٥, whereas the galactose ring interacts with the side chain of Trp٣٧٨. The side chain of Trp٣٨٦ constitutes an additional boundary to the active site cavity. There is no observable distortion of the G₃M₃ Glucose moiety from a chair conformation, as might have been expected in a Michaelis complex (49), suggesting that the side chain of the nucleophile Glu٣٥١ may be required for substrate distortion.

Both hydrocarbon chains of the G₃M₃ cereamide 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 Leu١٨٠, Tyr١٨٢, Ile١٨٣, Phe٢٣٥, Ile٢٧٦, and Leu٣٠٨. Upon binding of G₃M₃, the side chains of Arg١٧٧ and Asp٣١١ 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 Arg١٧٧.

EGC-Lactosyl-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. 1c). 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, Glu٣٥١ (Fig. 4).
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 GM3 and possess similar interactions with the enzyme. There are no discernable conformational differences between the EGC/H18528Lac and the apo-EGC structure. As with the apo structure, however, the ceramide-binding regions, in particular the side chain of Arg177 and the 308–317 loop, are poorly defined in the EGC/H18528Lac structure and have not been modeled. The identification of the glycosyl-enzyme intermediate confirms the identity of Glu351 and Glu233 as the catalytic nucleophile and general acid/base, respectively.

Attempts to soak EGC/E233A crystals with 2,4-dinitrophe-noyl/H9252-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...
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-G₃M₃ 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 catalytic center is lined with hydrophobic residues. Interestingly, the lipid-binding topology observed in the EGC-G₃M₃ 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 G₃M₃ activator protein (G₃M₃AP) (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-G₃M₃ 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 G₃M₃AP (~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 G₃M₃AP 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 \( \beta \)-hexosaminidase A. This is a family 20 glycosidase that catalyzes \( \text{G}_{\text{M2}} \text{AP} \) dependent \( \text{exo} \)-hydrolysis of the terminal \( \text{N} \)-acytetyl-\( \text{D} \)-galactosamine of \( \text{G}_{\text{M2}} \) to yield \( \text{G}_{\text{M3}} \). Predictive models of a \( \beta \)-hexosaminidase \( \text{A} \)-\( \text{G}_{\text{M2}} \text{AP} \)-\( \text{G}_{\text{M2}} \) complex based on mutagenesis analysis (65) suggest binding of only the sugar moiety by \( \beta \)-hexosaminidase \( \text{A} \), whereas the hydrophobic lipid tails remain bound by \( \text{G}_{\text{M2}} \text{AP} \). Hence, in comparison with EGC, a lipid binding site is not required by \( \beta \)-hexosaminidase \( \text{A} \), reflecting the significant differences in substrate binding imposed by \text{exo}- or \text{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 \text{d-erythro}-sphingosine was 40 \( \mu \text{M} \), which was increased by 10-fold for the hydrogenated analog \text{d-erythro}-dihydrosphingosine, although \( k_{\text{cat}} \) was roughly the same for both compounds, 0.1 \( \text{s}^{-1} \). When more substantial changes were introduced into the head group, the catalytic efficiency was dramatically reduced, making enzyme saturation impossible. When \text{l-threeo}-sphingosine and its hydrogenated analog \text{l-threeo}-dihydrosphingosine, in which both stereocenters of the head group are inverted relative to \text{d-erythro}-sphingosine, were employed, a 2000-fold reduction in \( k_{\text{cat}} / K_M \) 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 \text{d-erythro}-sphingosine, giving phytosphingosine, proved to be even more catastrophic, resulting in a 10,000-fold reduction in \( k_{\text{cat}} / K_M \).

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 \text{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 Henrias 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) Glycoconjug. 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. 227, 270–274
5. Horibata, Y., Okino, N., Ichinose, S., Omori, A., and Ito, M. (2000) J. Bioi. Chem. 275, 31297–31304
6. Horibata, Y., Sakaguchi, K., Okino, N., lida, H., Inagaki, M., Fujisawa, T., Hama, Y., and Ito, M. (2004) J. Bioi. Chem. 279, 33379–33389
7. Ito, M., and Yamagata, T. (1986) J. Bioi. 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. Bioi. 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. Bioi. Chem. 266, 7919–7926
15. Vaughan, M. D., Johnson, K., Defrees, S., Tang, X., Warren, R. A. J., and Withers, S. G. (2000) J. Am. Chem. Soc. 128, 6300–6301
16. Hakomori, S. (2000) Glycoconj. J. 17, 627–647
17. Nojiri, H., Yamana, H., Shirozu, G., Suzuki, T., and Isao, 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 Wilkelso, C. (2002) Dement. Geriatr. Cogn. 14, 128–136
21. Pope-Coleman, A., and Schneider, J. S. (1998) Rest. Neurol. Neuosci. 12, 255–266
22. Itoh, M., Fukumoto, S., Yamamoto, 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., Kurobe, Y., Sano, M., Kondo, A., Kato, I., and Ito, M. (1997) J. Bioi. Chem. 272, 19846–19850
26. Dvor, 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. Bioi. 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.,...
Structural Analyses of endo-Glycoceramidase II

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. (1993) 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. Vasella, A., Davies, G. J., and Bohm, M. (2002) Curr. Opin. Chem. Biol. 6, 619–629

47. Wright, C. S., Zhao, Q., and Rastinejad, F. (2003) J. Mol. Biol. 331, 951–964

48. 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