The Interaction of a Carbohydrate-binding Module from a *Clostridium perfringens* N-Acetyl-β-hexosaminidase with Its Carbohydrate Receptor*1,2

Elizabeth Ficko-Bleau1 and Alisdair B. Boraston2

From the Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

*Clostridium perfringens* is a notable colonizer of the human gastrointestinal tract. This bacterium is quite remarkable for a human pathogen by the number of glycoside hydrolases found in its genome. The modularity of these enzymes is striking as is the frequent occurrence of modules having amino acid sequence identity with family 32 carbohydrate-binding modules (CBMs), often referred to as F5/8 domains. Here we report the properties of family 32 CBMs from a *C. perfringens* N-acetyl-β-hexosaminidase. Macaray, UV difference, and isothermal titration calorimetry binding studies indicate a preference for the disaccharide LacNAc (β-D-galactosyl-1,4-β-D-N-acetylglucosamine). The molecular details of the interaction of this CBM with galactose, LacNAc, and the type II blood group H-trisaccharide are revealed by x-ray crystallographic studies at resolutions of 1.49, 2.4, and 2.3 Å, respectively.

*Clostridium perfringens* is a Gram-positive, spore-forming, non-motile, rod-shaped anaerobe. As a pathogen of humans, *C. perfringens* is often associated with gas gangrene, necrotic enteritis, and, most commonly, food poisoning (1–3). Determination of the genome sequence of *C. perfringens* (strain 13) (4) has revealed at least 54 open reading frames encoding putative glycoside hydrolases falling into 24 known glycoside hydrolase families (see afmb.cnrs-mrs.fr/CAZY/index.html) (5). Many encode intracellular proteins likely involved in the latter stages of sugar metabolism or proteins involved in peptidoglycan remodeling; however, roughly one-half are predicted to be secreted and are likely involved in the early stages of sugar metabolism. Although *C. perfringens* is most frequently thought of as a “flesh-eater,” its most common niche in humans is the gastrointestinal tract. However, very few, if any, of the secreted glycoside hydrolases have predicted substrate specificities consistent with metabolism of dietary polysaccharides in the human gut making gastric mucins, highly hydrated glycoproteins comprising up to 80% carbohydrate, the most likely target of the secreted *C. perfringens* enzymes. Indeed, the majority of these enzymes are predicted to have specificities appropriate for the degradation of complex glycans, suggesting that this bacterium is well equipped to attack the diverse sugar structures of the mucins in this environment. Consistent with this is the mucosal necrosis associated with severe enteritis caused by *C. perfringens* (6), which may be in part due to the arsenal of *C. perfringens* glycoside hydrolases. In turn, breaking down the mucosal barrier could improve access of other toxins, such as the pore-forming cpe (*C. perfringens* enterotoxin), to the epithelial layer.

Thirteen of the predicted *C. perfringens* (strain 13) glycoside hydrolases (and notably 13 glycoside hydrolases for each of the sequenced *Bacteroides* sp. genomes (*thetaiotaomicron*, *fragilis* YCH46, and *fragilis* 25285)) are highly modular and have, in addition to catalytic domains, modules with amino acid sequence identity to family 32 carbohydrate-binding modules (CBMs) (7). CBMs are generally considered to be modules with carbohydrate-binding function, but no catalytic activity, that are found within the modular architectures of glycoside hydrolases (8). They are currently classified into 45 families based on amino acid sequence identity (see afmb.cnrs-mrs.fr/~cazy/CAZY/index.html) and loosely grouped into three types, A (crystalline polysaccharide binding), B (polysaccharide chain binding), and C (small sugar binding or “lectin-like”), established by functional properties (8). Based on limited biochemical evidence, the family 32 CBMs appear to be type C CBMs. The x-ray crystal structures of the *Cladobteryum dendroides* galactose oxidase and the *Micromonospora viridifaciens* sialidase (MvGH33) revealed the β-sandwich lectin-like folds of their cognate CBM32 modules (9, 10). Co-crystallizations of MvGH33 with galactose showed the potential of its CBM32 module, here called MvCBM32, to bind galactose (10, 11), which was subsequently verified by functional studies (7). Thus, the family 32 CBMs, which are often referred to as F5/8 domains, have been generally considered as galactose binding domains.

The family 32 CBMs stand out among the CBM families, because they are frequently found appended to enzymes with

1 This work was supported in part by a grant from the Canadian Institutes for Health Research. 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.

2 The on-line version of this article (available at http://www.jbc.org) contains supplemental Fig. S1.

The atomic coordinates and structure factors (code 2j1a, 2j1e, and 2j1f) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

1 Recipient of Doctoral Fellowships from the Michael Smith Foundation for Health Research and the National Science and Engineering Research Council of Canada.

2 A Canada Research Chair in Molecular Interactions. To whom correspondence should be addressed: Biochemistry & Microbiology, University of Victoria, P. O. Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada. Tel.: 250-472-4168; Fax: 250-721-8855; E-mail: boraston@uvic.ca.

3 The abbreviations used are: CBM, carbohydrate-binding module; ITC, isothermal titration calorimetry; r.m.s.d., root mean square deviation.
“exotic” specificities (e.g. sialidases, β-hexosaminidases, mannosidases, and fucosidases) and are found in bacteria capable of causing disease in humans. In contrast, the vast majority of CBMs in other families are found appended to enzymes that are active on plant cell wall polysaccharides. In the context of the plant cell wall hydrolases, the function of CBMs has been repeatedly shown to be to localize the enzyme to an appropriate substrate (8). By analogy to the plant cell wall hydrolases, the role of family 32 CBMs is likely to target their parent enzymes to carbohydrate substrates; however, with these CBMs the substrates are likely more complex glycans, such as gastric mucins in the case of \( C. \) perfringens and Bacteroides sp. Previous studies of family 32 CBMs have not addressed the possibility that the specificity of these CBMs extend beyond a preference simply for galactose and may actually include specificity for complex glycan chains. Thus, studies of family 32 CBMs from bacterial pathogens enter a new area of carbohydrate-binding module-mediated host-pathogen interactions and will extend our knowledge of this potentially complex family of carbohydrate-binding proteins.

**EXPERIMENTAL PROCEDURES**

**Materials**—Unless otherwise stated, chemicals, carbohydrates, glycoproteins, and polysaccharides were purchased from Sigma.

**Cloning**—The DNA fragment encoding the family 32 CBM (Fig. 1) of \( C. \) perfringens was amplified by PCR from \( C. \) perfringens genomic DNA (Sigma, ATCC 13124) using previously described methods (15). Nucleotides 1873–2301 of the \( \text{cpgh84c} \) gene, corresponding to the CBM (amino acid residues 625–767), were amplified with the oligonucleotide primers 5′/H11032-CACCAATCCAA-GAACAGTAAAG-3′ (CBMF) and 5′/H11032-CTTTTATCCATGAAC-ATTAACCTC-3′ (CBMR). The amplified gene fragment was ligated directly into the pET-150 TOPO Directional Cloning kit (Invitrogen) to generate pCBM32. The polypeptide (called \( \text{CpCBM32} \)) encoded by pCBM32 comprises a H6 tag fused to the \( \text{CpCBM32} \) module by an enterokinase protease cleavage site.

To better understand CBM32 structure and function we initiated studies of \( \text{CpGH84C} \) from \( C. \) perfringens (strain ATCC 13124). This enzyme, which comprises four modules defined on the basis of primary structure comparisons (see Fig. 1), was chosen as a model system, because, relative to other family 32 CBM-containing enzymes, it is reasonably small and has a simple modular architecture that is amenable to accurate definition of the modular boundaries. To facilitate structure-function studies, we dissected this protein at the genetic level to recombinantly produce isolated \( \text{CpCBM32} \). The experimental results reveal the ability of the CBM to bind to terminal glycotopes commonly found in elaborated O- and complex N-glycans (12–14), whereas the x-ray crystal structures of \( \text{CpCBM32} \) in complex with sugar help uncover the molecular details that confer this binding ability. Structural comparisons with known CBM32s and other \( C. \) perfringens CBM32s suggest variations in glycan specificities, but all are based on a key terminal galactose residue. This work provides the first detailed structure-function analysis of a family 32 CBM and will provide a foundation for further studies of CBMs within this family.
Characterization of a Family 32 CBM

TABLE 1

| X-ray data collection and refinement statistics | Galactose | LacNAc | H-trisaccharide |
|------------------------------------------------|-----------|--------|-----------------|
| Space group                                      | P32,2     | P42,2  | P42,2           |
| Unit cell                                        | a = 43.84, b = 43.84, c = 139.47; | α = 90.00, β = 90.00, γ = 120.00; | α = 90.00, β = 90.00, γ = 120.00; |
| Asymmetric unit contents                         | Monomer   | Monomer | Monomer         |
| Resolution range                                 | 20.1-49 (1.53-1.49) | 20.240 (2.49-2.40) | 20.230 (2.36-2.30) |
| Number of measured reflections                   | 23,619    | 31,226 | 18,519          |
| Number of unique reflections                     |           |        |                 |
| Redundancy (%)                                   | 67.0 (3.11) | 67.0 (2.7) | 67.0 (2.2)      |
| Completeness (%)                                 | 20.6      | 25.3   | 25.3            |
| Rmerge (%)                                       | 0.018     | 0.019  | 0.019           |
| Rvalue (%)                                       | 27.79     | 25.2   | 24.0            |
| Rfree (%)                                        | 0.108     | 0.107  | 0.088           |
| r.m.s. bond lengths (Å)                          | 27.79     | 25.2   | 24.0            |
| r.m.s. chiral-centre restraints (Å²)              | 0.108     | 0.107  | 0.088           |
| All atoms                                        | 22.40     | 27.79  | 30.67           |
| Protein atoms                                    | 20.88     | 26.57  | 29.08           |
| Water molecules                                  | 32.24     | 34.79  | 36.66           |
| Sugar                                            | 16.38     | 29.13  | 45.64           |
| No. residues                                     | 143       | 142    | 143             |
| No. waters                                       | 172       | 143    | 218             |
| No. sugar atoms                                  | 12        | 26     | 36              |
| PDB code                                         | 2j1a      | 2j1e   | 2j1f            |

1-thio-β-D-galactopyranoside then grown overnight at 37 °C. The cells were harvested at 4,000 × g and resuspended in 20 ml of binding buffer containing 20 mM Tris, pH 8, and 0.5 M NaCl. Cells were lysed using a French pressure cell. Cell debris was removed by centrifugation for 1 h at 27,000 × g. The supernatant was applied to His-Select resin followed by step elution with binding buffer containing imidazole concentrations between 5 and 500 mM. Samples were run on a 15% SDS polyacrylamide gel, and fractions containing the polypeptide of interest were pooled. Proteins were concentrated, and buffer exchange was done into phos-
Characterization of a Family 32 CBM

Diffraction data were collected with a Rigaku R-AXIS IV++ area detector coupled to an MM-002 x-ray generator with Osmic “blue” optics and an Oxford Cryostream 700. Data were processed with Crystal Clear/d’trek (21). All data collection statistics are given in Table 1.

Structure Determination—CpCBM32 was solved by molecular replacement using the family 32 galactose-binding module from the M. viridifaciens sialidase (pdb code 1EUT) (10) as a search model. The program molrep (22) was able to find one clear rotation/translation solution corresponding to the single molecule in the asymmetric unit. This initial model was corrected, and ligand was added by successive rounds of building using COOT (23). Refinement was done using REFMAC (24). Water molecules were added using the REFMAC implementation of ARP/wARP and inspected visually prior to deposition. The final model lacking waters was used as a starting model to solve the structures of the other CpCBM32 sugar complexes. Initial models were corrected, one or more ligands were added, and waters were added as above. Residue numbering conforms to the numbering in the complete CpGH84C enzyme. Final model statistics are given in Table 1.

RESULTS

Carbohydrate Binding Properties of CpCBM32 from CpGH84C—Based on its amino acid sequence identity (~30%) with the family 32 galactose-binding module from the M. viridifaciens sialidase we postulated that the CpCBM32 module in CpGH84C is indeed a carbohydrate-binding protein. This was initially investigated by macroarray binding experiments using arrayed glycoproteins and polysaccharides, which revealed significant binding to asialofetuin, type III porcine gastric mucin, and fetuin with relative binding of asialofetuin > porcine gastric mucin > fetuin (supplemental Fig. S1). Using knowledge of the glycan structures commonly found on these glycoproteins (14) as a guide, we assessed the binding of various monosaccharides, disaccharides, and trisaccharides to CpCBM32 by qualitative and quantitative UV difference experiments. The addition of d-galactose and GalNAc (N-acetyl-d-galactosamine) to CpCBM32 resulted in large perturbations of the UV difference spectra indicative of the involvement of tryptophan in sugar binding (17) (Fig. 2A). L-Fucose, d-glucose, d-mannose, and GlcNAc (N-acetyl-d-glucosamine) were also tested but did not influence the UV absorption of CpCBM32 and thus are unlikely primary ligands of CpCBM32. Quantitative studies by UV difference titrations showed an affinity of roughly $1 \times 10^3$ M$^{-1}$ for galactose-based monosaccharides and a preference

Crystallization and Data Collection—All crystallization experiments were performed using the hanging-drop vapor-diffusion method. Prior to crystallization, the H6 tag was removed from CpCBM32 by treatment with enterokinase over a 4-day period. The digested sample was run through a NovaGen His-bind Quick 900 cartridge to remove the His tag and any undigested protein from the solution. Samples were concentrated and exchanged as above into 20 mM Tris, pH 8. Co-crystals of CpCBM32 (10.5 mg/ml) with galactose (~10 mM) were obtained with 0.2 mM MgCl$_2$, 25% polyethylene glycol 2000 monomethylether, and 0.1 mM Tris, pH 7.5. These crystals were cryoprotected with 15% glycerol in mother liquor. 1.5 mM sodium/potassium phosphate was used to co-crystallize CpCBM32 (20 mg/ml) LacNAc and the type II blood group H-trisaccharide (α1-2-fucosyl-1,2-β-d-galactosyl-1,4-β-d-N-acetylglucosamine). Optimization of this condition determined the ideal Na$_2$HPO$_4$/K$_2$HPO$_4$ concentration to be 1.5 mM (at a 1:100 ratio) and the optimal protein concentration to be 5 mg/ml. The cryoprotectant used was 1.45 mM Na/ K$_2$HPO$_4$ with 27% ethylene glycol.

TABLE 2

| Sugar            | $K_a$ (M$^{-1}$) |
|------------------|-----------------|
| d-Galactose      | 0.98 (±0.17)    |
| d-Glucose        | 0.86 (±0.12)    |
| Methyl-α-d-galactose | 0.57 (±0.25)  |
| Methyl-β-d-galactose | 1.31 (±0.08)  |
| Lactose          | 2.49 (±0.06)    |
| LacNAc           | 9.09 (±2.98)    |

Figure 2. Assessment of CpCBM32 binding. A, UV difference spectrum produced by the addition of excess galactose to CpCBM32 (60 μM). Wavelengths at which peaks and troughs occur are labeled. B, an example of UV difference binding isotherm produced by titrating lactose into CpCBM32 (45 μM) in 50 mM Tris-HCl, pH 7.5. The solid line shows the fit to a one-binding site model.
Characterization of a Family 32 CBM

TABLE 3

Binding parameters determined by ITC in 50 mM Tris-HCl, pH 7.5

| Sugar          | Temperature | n   | $K_a$  | $\Delta H$ | $\Delta S$ | $\Delta G$ |
|----------------|-------------|-----|--------|------------|------------|------------|
|                | °C          | n  | $10^{-4}$ | kcal/mol   | cal/mol/K  | kcal/mol   |
| LacNAc         | 25          | 0.89 ± 0.0 | 1.11 (±0.02) | −15.6 (±0.0) | −33.6 (±0.5) | −5.5 (±0.0) |
|                | 20°         | 0.89 ± 0.0 | 1.77 (±0.02) | −15.1 (±0.1) | −32.1 (±0.5) | −5.7 (±0.0) |
|                | 17.5°       | 0.89 ± 0.0 | 2.26 (±0.02) | −14.5 (±0.1) | −30.8 (±0.5) | −5.8 (±0.0) |
| H-trisaccharide| 25°         | 1.0° | 0.12 (±0.00) | −6.6 (±0.1)  | −8.3 (±0.1)  | −4.2 (±0.0)  |

* Only single titrations were performed, and errors are those from the data-fitting process. Otherwise, errors are the standard deviations determined from three independent experiments.

* This value was fixed as a constant for the data-fitting process.

FIGURE 3. Structural features of CpCBM32. A, secondary structure stereo-view representation of CpCBM32. The tryptophan in the carbohydrate binding site and the bound galactose are shown in gray stick representation. B, coordination of the bound metal ion. Only the side chains involved in binding the metal are shown for clarity. In both panels the metal ion is shown in green.

for β- rather than α-configured O-methylgalactose (Fig. 2B and Table 2). The presence of the acetamido group of GalNAc did not appear to confer any advantage to binding. CpCBM32 preferred lactose and GalNAc over galactose by factors of ~2.5- and 10-fold, respectively. The increased affinity of LacNAc (cf. lactose) suggested the specific involvement of the 2'-acetamido group of the GlcNAc moiety in binding. Along similar lines, the α-1,2-linked 1-fucose of fucosylactose (α-1-fucosyl-1,2-β-D-galactosyl-1,4-β-D-glucose) substantially reduced the binding affinity to levels below those we could quantify due to limiting quantities of sugar but did not entirely legislate against binding.

The binding to LacNAc and the type II H-trisaccharide was further investigated by ITC (Fig. 2, C and D, and Table 3). The values for LacNAc revealed the enthalpically driven binding process common to protein-carbohydrate interactions (25–27). The $\Delta H$ of binding was temperature-dependent, the analysis of which allowed the approximation of the change in heat capacity ($\Delta C_p$) to be −105 (±5) cal/mol/K. Again, this small negative $\Delta C_p$ is consistent with the majority of protein-carbohydrate interactions. The affinity of CpCBM32 for the type II H-trisaccharide was too low to accurately deconvolute the stoichiometry of binding, so, on the basis of the LacNAc binding data and x-ray crystallography data (see below), this value was fixed at 1 for the analysis (28). Like with fucosylactose, the fucose residue of the type II H-trisaccharide is detrimental to binding relative to LacNAc or lactose, although it does not destroy binding. The roughly 1.3 kcal/mol increase in free energy due to the fucose moiety on the H-trisaccharide appears to come by virtue of a substantial enthalpic penalty (~+9 kcal/mol), which is partially offset by a favorable contribution to entropy (~+25 cal/mol/K) (note: assumption of up to a 25% error in the estimate of stoichiometry due to errors in either the sugar or protein concentration does change the magnitudes of the calculated thermodynamic penalties but does not change their sign and, thus, qualitative interpretation of the data are unaffected).

The Fold of CpCBM32—CpCBM32 was crystallized in the presence of galactose, and its x-ray crystal structure was solved by molecular replacement at high resolution (1.49 Å). The fold is that of a β-sheet comprising three anti-parallel β-strands opposing a β-sheet of five anti-parallel β-strands (Fig. 3A). The closest structural neighbors are the family 32 CBMs from the fungal Cladobotryum dendroides galactose oxidase (PDB code 1GOF (9)) (here the CBM is referred to as CdCBM32) and the bacterial M. viridifaciens sialidase (PDB code 1EUT (10)) (here the CBM is referred to as MvCBM32; see below for a more detailed comparison). More distantly related are the family 6 and 36 CBMs as well as the Anguilla sp. fucolactin.

CpCBM32 heptahedrally coordinates one metal ion through the side chains of Thr-655, Asp-650, and Glu-672 and the backbone carbonyls of Phe-647, Asp-652, Thr-655, and Ala-761 (Fig. 3C). This ion was judged most likely to be calcium on the basis of three observations. Firstly, the atom possessed significant anomalous scattering properties, because it could be easily located in anomalous difference maps, indicating that the ion was not sodium or magnesium. Secondly, when considering ions commonly found associated with carbohydrate-binding proteins, heptahedral coordination mediated entirely by oxygen atoms is most consistent with potassium, calcium, or manganese. Lastly, when modeled as calcium, the B factor of this
atom was consistent with the B factors of neighboring atoms in the coordinating side chains. This ion does not appear to play a direct role in binding carbohydrate, because it is quite distant.

FIGURE 4. Representative electron density and interaction of CpCBM32 with galactose (A), LacNAc (B), and the type II blood group H-trisaccharide (C). All maps are maximum-likelihood/σA (37)-weighted 2F_{obs} − F_{calc} electron density maps contoured at 1σ (0.35, 0.30, and 0.29 electrons/Å³ for galactose, LacNAc, and the H-trisaccharide, respectively). The disordered loop in the galactose complex is shown with a dashed line. Amino acid side chains involved in binding the sugars are shown in gray stick representation and labeled.

FIGURE 5. Schematics showing the interactions of CpCBM32 with galactose (A), LacNAc (B), and the type II blood group H-trisaccharide (C). A distance of 3.2 Å was used as the cut-off for determination of significant hydrogen bonds. Water molecules are shown as shaded spheres.

Characterization of a Family 32 CBM
Characterization of a Family 32 CBM

A

B

C
from the carbohydrate binding site. Studies on other CBMs where similar “structurally” relevant ions have been removed have suggested only a stabilizing role for such ions, because apo-CBMs are still competent to bind carbohydrate (29, 30). It is notable that the structural position of this ion is also conserved in CBMs from family 6 (7, 31, 32), 36 (33), and the Anguilla sp. fucolectins (34).

**CpCBM32 in Complex with Carbohydrates**—The electron density map of the galactose complex clearly revealed a single bound molecule of galactose. Trp-661 and Phe-757 create a relatively hydrophobic pocket, which cradles the C6-hydroxymethyl group. Trp-661 “stacks” against the flat, apolar surface created by carbons 3–6 on the B-face of β-galactose with the O4 pointing away from the aromatic residue (Fig. 4). Specificity for the non-reducing end of β-galactose (and presumably GalNAc) is conferred by three potential hydrogen bonds to the axial O4 of the sugar from the terminal δO of Asn-695, a terminal guanido nitrogen of Arg-690, and the εN from the imidazole ring of His-658 (Fig. 5). Additional hydrogen bonds are made between the O3 of β-galactose and Arg-690 and Glu-641; the endocyclic oxygen of galactose makes a hydrogen bond with the amide nitrogen of Asn-695. The burial of the galactose O3 would prevent binding to the AB blood group antigens.

Examination of the CBM32–galactose complex revealed that the O1, O2, and O6 groups of the bound galactose were solvent-exposed, thus hinting at how this protein might accommodate the additional sugar groups of LacNAc and the type II H-trisaccharide. We probed the presence of additional subsites by co-crystallizing CpCBM32 with these possible biological ligands of CpCBM32 (Fig. 4). The binding of LacNAc appears to induce some changes relative to the galactose structure. In the galactose structure the loop comprising residues 750–753 was very disordered, and, in fact, residues 751 and 752 could not be modeled. In the LacNAc structure, this loop becomes ordered, and the side chains of Asp-749 and Gln-750 close in on the sugar to make additional hydrogen bonds with both the galactose and GlcNAc residues (Figs. 4 and 5). One terminal δO of Asp-749 is situated to hydrogen bond with the O3 of GlcNAc. The other terminal δO of Asp-749 along with the δO of Ser-693 positions an ordered water for a water-mediated hydrogen bond to the oxygen of the β1,4-linked glycosidic bond. The terminal amide nitrogen of Gln-750 is positioned to hydrogen bond with the O6 of galactose and the O3 of GlcNAc. The 2′-acetamido group of the GlcNAc does not appear to be positioned to make any additional hydrogen bonds. However, this chemical group does sit above a relatively apolar platform made by the planar conformations of the Asp-749 and Gln-750 side chains and makes a number of van der Waals contacts. A water-mediated hydrogen bond occurs between the carbonyl group oxygen of GlcNAc, and the main-chain nitrogen of Ala-751. These additional interactions are the likely source of the increased affinity of CpCBM32 for LacNAc versus lactose.

The crystal structure of CpCBM32 in complex with the type II blood group H-trisaccharide revealed the interactions between the LacNAc core of this sugar and the protein to be identical to the LacNAc-CpCBM32 interactions (Fig. 4). The fucose residue of the type II H-trisaccharide occupied a subsite located above Arg-690, and hydrogen bonding interactions with the protein were limited to two water-mediated hydrogen bonds: the first provided by the δO of Asp-749 and the δO of Ser-693 (Fig. 5). The same water is also involved in a water-mediated hydrogen bond to the glycosidic bond between galactose and GlcNAc. The second water is positioned for potential hydrogen bonding with the O2 of fucose by the εO of Glu-641.

**DISCUSSION**

**Molecular Determinants of CpCBM32 Specificity**—The CBM32 module from CpGH84C appears to have three, possibly four, subsites that accommodate the monosaccharide units of oligosaccharide ligands. The primary subsite binds galactose but is also able to accommodate GalNAc, and it appears to provide the bulk of the binding free energy and thus provide the “anchor” for the interaction. β-Linked substituents on the O1 of this galactose result in improved binding affinity. In the case of LacNAc, where the GlcNAc is β-1,4-linked to the galactose, the affinity improves by an order of magnitude relative to galactose and is influenced positively by the presence of the acetamido group of GlcNAc. Modeling of lacto-N-biose (β-D-galactosyl-1,3-β-D-N-acetylgalcosamine; a component of the type I blood group H-trisaccharide antigen) into the CpCBM32 binding site, using LacNAc as a guide, suggested that CpCBM32 may also accommodate this sugar with the β-1,3-linked GlcNAc bound in the same secondary subsite as for the GlcNAc of LacNAc (not shown). In the case of lacto-N-biose, the C6-hydroxymethyl group of the GlcNAc residue in this disaccharide would be positioned similarly to the acetamido group of LacNAc. Although structurally there is apparently nothing to hinder the binding of lacto-N-biose, it is unknown if the C6-hydroxymethyl interaction of this sugar
Characterization of a Family 32 CBM

would be energetically equivalent to those of the LacNAc acetamido group. Along an identical line of argument, the core 1 O-glycan (β-d-galactosyl-1,3-β-d-N-acetylglactosamine) would be accommodated in a similar manner to lacto-N-biose.

The structure of CpCBM32 with the type II H-trisaccharide revealed an additional secondary subsite that accommodates the β-1,2-linked fucosyl residue. However, this appeared to incur a substantial energetic penalty. The thermodynamic values indicated an enthalpically unfavorable but partially offsetting entropically favorable contribution from the occupation of this subsite. This signature is consistent with the removal of ordered waters from the protein surface. Indeed, when comparing the binding site water networks in the d-galactose, LacNAc, and type II H-trisaccharide structures, the latter sugar results in the displacement of at least one, possibly two, waters that would otherwise be ordered on the protein surface in the absence of the fucosyl residue. Overall, although the fucosyl residue of the blood group H-antigen is accommodated by the CBM, it is not a preferred feature.

CpCBM32 may have a fourth subsite that accommodates substitutions linked to the O6 of the anchoring galactose. The O6 of galactose is solvent-exposed, which may permit the linkage of another sugar to that atom. Sialic acid is found linked 2–6 to LacNAc, typically at the terminus of a glycan, and to GalNAc, often as an elaboration of O-glycan cores (14). It is possible that CpCBM32 may accommodate a sialic acid residue within this putative fourth binding subsite.

Comparison with Other Family 32 CBMs—The closest structural neighbors of CpCBM32 are MvCBM32 and CdCBM32 with amino acid sequence identities of 35 and 30%, respectively, and r.m.s.d. values for both of 1.2 Å² over 135 and 130 matched Cα atoms, respectively (Fig. 6, A and B). Clearly, the backbones of CpCBM32, MvCBM32 and CdCBM32 match very closely with the exception of a single loop in CpCBM32 (Fig. 6, A and B). Recently, the high resolution crystal structure of MvCBM32 in complex with galactose was solved (11). Although a similar complex for CdCBM32 is not available, a comparison of the MvCBM32 and CdCBM32 structures (not shown) and the residues involved in galactose binding show excellent conservation of functional residues (Fig. 6A). However, in the absence of CdCBM32 complex, further comparison is uninformative. A comparison of MvCBM32 and CpCBM32 reveals that the architecture of the carbohydrate-binding sites and the constellation of protein–carbohydrate interactions are very similar. The majority of the galactose binding machinery comprising Glu-641, His-658, Trp-661, Arg-690, and Phe-757 of CpCBM32 is extremely well conserved between the two proteins (Fig. 6, A and C). Additionally, in binding galactose CpCBM32 employs Asn-695 whose structural analog in MvCBM32 is not positioned to interact with the galactose. The loop comprising residues 747–752 in CpCBM32 (indicated by the green bar in Fig. 6A and the arrows in Fig. 6, B and C) harbors the additional residues Asp-749 and Gln-750, which appear key in binding the GlcNAc residue of LacNAc. That MvCBM32 (and CdCBM32) lacks this loop suggests that LacNAc is not the primary ligand for these CBMs.

This galactose recognition machinery is well conserved among several other clostridial CBM32s despite having amino acid sequence identity as low as 25% (Fig. 6A). These enzymes span specificities for α- or β-linked glucose, GlcNAc, galactose, GalNAc, and sialic acid sugars. The highest degree of sequence variability occurs in the regions of the polypeptides that contribute to the proposed additional binding subsites (Fig. 6A). This suggests that, among these reasonably closely related CBM32s, galactose is the primary ligand but that different substitutions on the galactose may be optimally accommodated by the different CBMs. This added diversity could complement the varied specificities of the enzymes to which these CBM32 modules are appended. However, precisely how CBM32 ligand preference might complement the enzyme specificity is unclear. For example, the receptor specificity of the CpCBM32 from CpGH84C for terminal LacNAc motifs does not match the specificity of CpGH84C, which evidence indicates is an exo-β-d-N-acetylglucosaminidase (15, 35, 36).

Using amino acid sequence identity criteria less stringent than a 25% cut-off allows the identification of numerous additional family 32 CBMs in C. perfringens and other bacteria. However, at these lower levels of sequence identity the conservation of functional residues, including those involved in galactose binding, is lost, and only amino acids contributing to structure remain conserved (not shown). This suggests that the specificity of family 32 CBMs is not necessarily centered on galactose.

CONCLUSION

Roughly one-third of the bacterial enzymes containing family 32 CBMs belong to a varied range of glycoside hydrolases produced by the bacterial gut colonizers from species of Clostridium and Bacteroides. The role of these enzymes is likely in scavenging carbohydrates from dietary sugars or muco-oligosaccharides. By analogy to plant cell wall hydrolases, the role of the CBM32s in these enzymes is probably to attach the enzyme to a carbohydrate-bearing surface, which in the case of CpGH84C is likely the terminal LacNAc motif common to the O-linked glycans of mucin. This evokes an image of the enzyme being attached to a carbohydrate-bearing surface via the CBM while allowing the catalytic module to “graze” on that substrate. Given the diversity of enzymes that contain family 32 carbohydrate binding modules, one might expect, like with plant cell wall hydrolases and their CBMs, that the diversity of catalytic specificities is mirrored by the diversity of CBM32 specificities. Thus, CBM family 32 may be a resource to be mined for carbohydrate-binding proteins with a variety of specificity profiles.

REFERENCES

1. Hatheway, C. L. (1990) Clin. Microbiol. Rev. 3, 66–98
2. McDonel, J. L. (1980) Pharmacol. Ther. 10, 617–655
3. Rood, J. I., and Cole, S. T. (1991) Microbiol. Rev. 55, 621–648
4. Shimizu, T., Ohshima, S., Ohtani, K., and Hayashi, H. (2001) Microbiol. Immunol. 45, 179–189
5. Coutinho, P. M., and Henrissat, B. (1999) in Recent Advances in Carbohydrate Bioengineering (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B.), pp. 3–12, The Royal Society of Chemistry, Cambridge
6. Gerding, D. N., and Johnson, S. (1997) in The Clostridia Molecular Biology and Pathogenesis (Rood, J. I., McClane, B. A., Songer, J. G., and Titball, R. W.), pp. 117–140, Harcourt Brace & Company, London
7. Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R., and Davies, G. (2003) J. Mol. Biol. 327, 659–669
8. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Biochem. J. 382, 769–781
9. Ito, N., Phillips, S. E., Stevens, C., Ogé, Z. B., McPherson, M. J., Keen, J. N., Yadav, K. D., and Knowles, P. F. (1991) Nature 350, 87–90
10. Gaskell, A., Crennell, S., and Taylor, G. (1995) Structure 3, 1197–1205
11. Newstead, S. L., Watson, J. N., Bennet, A. J., and Taylor, G. (2005) Acta Crystallogr. D Biol. Crystallogr. 61, 1483–1491
12. Gupta, D., Kaltner, H., Dong, X., Gabius, H. J., and Brewer, C. F. (1996) Glycobiology 6, 843–849
13. Morelle, W., Haslam, S. M., Ziak, M., Roth, J., Morris, H. R., and Dell, A. (2000) Glycobiology 10, 295–304
14. Robbe, C., Capon, C., Coddeville, B., and Michalski, J. C. (2004) Biochem. J. 384, 307–316
15. Ficko-Blean, E., and Boraston, A. B. (2005) Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 61, 751–752
16. Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) Anal. Biochem. 200, 74–80
17. Boraston, A. B., Warren, R. A., and Kilburn, D. G. (2001) Biochemistry 40, 14679–14685
18. Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A., and Kilburn, D. G. (2001) Biochemistry 40, 6240–6247
19. Lammerts van Bueren, A., Finn, R., Ausio, J., and Boraston, A. B. (2004) Biochemistry 43, 15633–15642
20. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Anal. Biochem. 179, 131–137
21. Pflegrath, J. W. (1999) Acta Crystallogr. D Biol. Crystallogr. 55, 1718–1725
22. Vagin, A., and Teplyakov, A. (1997) J. Appl. Crystallogr. 30, 1022–1025
23. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
24. Murshed, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. D Biol. Crystallogr. 53, 240–255
25. Dam, T. K., and Brewer, C. F. (2002) Chem. Rev. 102, 387–429
26. Sultan, N. A., and Swamy, M. J. (2005) Arch. Biochem. Biophys. 437, 115–125
27. Bachhawat-Sikder, K., Thomas, C. J., and Surolia, A. (2001) FEBS Lett. 500, 75–79
28. Turnbull, W. B., and Daranas, A. H. (2003) J. Am. Chem. Soc. 125, 14859–14866
29. Rosse, Y., Sunna, A., Pfeil, W., and Heinemann, U. (2004) J. Mol. Biol. 340, 543–554
30. Johnson, P. E., Creagh, A. L., Brun, E., Joe, K., Tomme, P., Haynes, C. A., and McIntosh, L. P. (1998) Biochemistry 37, 12772–12781
31. van Bueren, A. L., Morland, C., Gilbert, H. J., and Boraston, A. B. (2005) J. Biol. Chem. 280, 530–537
32. Lammerts van Bueren, A., and Boraston, A. B. (2004) J. Mol. Biol. 340, 869–879
33. Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M., and Davies, G. J. (2004) Structure 12, 1177–1187
34. Bianchet, M. A., Odom, E. W., Vasta, G. R., and Amzel, L. M. (2002) Nat. Struct. Mol. Biol. 9, 628–634
35. Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K. A., Turkenburg, J. P., Hart, S. J., Black, G. N., Vocadlo, D. J., and Davies, G. J. (2006) Nat. Struct. Mol. Biol. 13, 365–371
36. Rao, F. V., Dorfmueller, H. C., Villa, F., Allwood, M., Eggleston, I. M., and van Aalten, D. M. (2006) EMBO J. 25, 1569–1578
37. Read, R. J. (1986) Acta Crystallogr. Sect. A 42, 140–149