Carbohydrate Recognition by a Large Sialidase Toxin from *Clostridium perfringens*[†]

Alisdair B. Boraston,* Elizabeth Ficko-Blean, and Michael Healey[‡]

Biochemistry & Microbiology, University of Victoria, P.O. Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada

Received July 4, 2007; Revised Manuscript Received July 25, 2007

ABSTRACT: Myonecrotic isolates of *Clostridium perfringens* secrete multimodular sialidases, often termed "large sialidases", that contribute to the virulence of this bacterium. NanJ is the largest of the two secreted sialidases at 1173 amino acids and comprises 6 different modules which are, from the N-terminus, a family 32 carbohydrate binding module (CBM), a family 40 CBM, a family 33 glycoside hydrolase, a module of unknown function, a family 82 "X-module" of unknown function, and a module with amino acid similarity to fibronectin type III domains. The hydrolase activity of clostridial sialidases is quite well documented; however, the functions of their accessory domains are entirely uninvestigated. Here we describe the carbohydrate binding activity of the isolated family 32 CBM (CBM32) and the isolated family 40 CBM (CBM40). CBM32 is shown to bind galactose or *N*-acetylgalactosamine, while CBM40 is sialic acid specific, though both CBMs appear to bind with very low affinities. The crystal structure of CBM32 was determined at 2.25 Å in complex with galactose. This revealed what appears to be a very simple galactose binding site. The crystal structure of CBM40 was determined at 2.20 Å in complex with a sialic acid containing molecule that it fortuitously crystallized with, revealing the molecular details of the CBM40−sialic acid interaction. Overall, the results indicate that NanJ contains carbohydrate specific binding modules that likely function to target the enzyme to molecules or cells bearing mixed populations of glycans that terminate in either galactose/*N*-acetylgalactosamine or sialic acid.

Clostridium perfringens is a ubiquitous Gram-positive, spore-forming, rod-shaped anaerobe. As a species, it is a prolific producer of toxins and secreted virulence factors (*1*, *2*). *C. perfringens* is biotyped into 5 groups (A−E) on the basis of differential production of four major toxins: α, *β*, *ϵ*, and *ι* (*1*, *2*). These proteins are classified as major toxins because they are capable of killing mice when injected intraperitoneally. The most clinically relevant biotype is type A, which produces the α-toxin (phospholipase C). It is one of the most common causes of food poisoning and a cause of gas gangrene in humans. More recently, this biotype has become a problem in the poultry industry (*3*). Reduced use of antibiotics in poultry farming has resulted in the increased occurrence of *C. perfringens* in these birds. The necrotic form has caused higher mortality rates, while a subclinical form of *C. perfringens* infection has resulted in reduced weight gain (*3*). Not only has the presence of *C. perfringens* in poultry flocks had an impact on yields, but it has also substantially increased the risk of transmission to humans (*3*).

In addition to the major toxins, *C. perfringens* produces a battery of toxins that are classified as minor due to their inability to cause lethal effects in mice. Among these minor toxins are the sialidases (*4*−*8*). These enzymes remove terminal sialic acids from glycans, which has two important outcomes during the infection of humans. First, these enzymes are known to potentiate the activity of the α-toxin, though the mechanism of this synergism is currently unknown (*9*). Also, clostridial sialidases are thought to activate the normally sialylated Thomsen−Friedenreich antigen by desialylating it (*10*−*12*). This is believed to cause life-threatening hemolysis during blood transfusions of *C. perfringens* infected individuals (*10*−*12*).

Depending on the strain, the *C. perfringens* genome can contain up to three different sialidase genes (*13*, *14*). Though the catalytic modules of all of these clostridial sialidases show amino acid sequence identity and belong to family 33 of the glycoside hydrolase classification (*15*), they have different modular structures. For example, the genomes of the two sequenced "flesh-eating" *C. perfringens* type A strains (strain 13 and ATCC 13124, strains isolated from gangrenous lesions) both contain genes encoding two large secreted sialidases, NanI and NanJ. NanJ has a complex multimodular architecture comprising a central catalytic module and five accessory modules. None of the accessory modules have known functions, but the N-terminal two modules show amino acid sequence identity with family 32 and family 40 carbohydrate binding modules (CBMs),[1] respectively (see below) (Figure 1). NanI is somewhat simpler, having only a

[†] This work was supported by a grant from the Canadian Institutes for Health Research. A.B.B. is a Canada Research Chair in Molecular Interactions. E.F-B. is supported by doctoral fellowships from the Natural Sciences and Engineering Research Council of Canada and the Michael Smith Foundation for Health Research.

* To whom correspondence should be addressed. Phone: (250) 472-4168. Fax: (250) 721-8855. E-mail: boraston@uvic.ca.

[‡] Current address: Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, BC V5A 1S6, Canada.

[1] Abbreviations: CBM, carbohydrate binding module; galNAc, *N*-acetylgalactosamine; IMAC, immobilized metal ion affinity chromatography; IPTG, isopropyl *β*-D-thiogalactopyranoside; ITC, isothermal titration calorimetry; K_a, association constant.

10.1021/bi701317g CCC: $37.00
Published on Web 09/13/2007

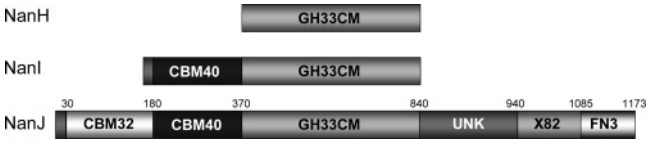

FIGURE 1: Modular organization of the clostridial sialidases. Amino acid numbers corresponding to the module boundaries are shown above the schematic of NanJ. The nomenclature is as follows: CBM32, family 32 carbohydrate binding module; CBM40, a module having amino acid sequence identity with family 40 CBMs; GH33CM, family 33 glycoside hydrolase catalytic module; UNK, modules having unknown function and little sequence identity with other proteins of known function; X82, modules with unknown function classified as family 82 X-modules; FN3, a module sharing distant identity with fibronectin type III domains.

catalytic module and an N-terminal family 40 CBM (Figure 1). The sequenced food poisoning isolate of *C. perfringens* (SM101) harbors a gene encoding a small intracellular sialidase, NanH, that comprises only a catalytic domain (ATCC 13124 also has this gene), but SM101 does not have genes for NanI or NanJ. Thus, the complex modularity of the *C. perfringens* sialidases appears to correlate with the myonecrotic abilities of the strain. The role of sialidase modularity in the pathogenesis of myonecrotic *C. perfringens* is currently unknown.

CBMs are contiguous amino acid sequences of ~40−200 residues in length that are commonly found in carbohydrate-active enzymes, such as glycoside hydrolases (*16*). These modules, which are currently classified into 49 families on the basis of amino acid sequence similarity, have only carbohydrate binding activity (i.e., no catalytic activity) and function to target the parent enzyme to relevant glycans (*16*). Though CBM research began with the study of CBMs involved in plant cell wall polysaccharide hydrolysis, CBMs are now frequently being found in carbohydrate-active enzymes from bacterial pathogens. Indeed, recent studies have indicated that targeting the adherent activity of CBMs in pathogenesis-related glycoside hydrolases with multivalent ligands/substrates may be a viable approach to developing carbohydrate-based therapeutics that inhibit these toxins (*17*). Family 32 and 40 CBMs, such as those found in the *C. perfringens* sialidases, are among the most frequently found types of CBMs in bacterial virulence factors. CBM32's appear to be primarily galactose specific (*18*−*21*), whereas the only characterized member of CBM40, that from the *Vibrio cholerae* sialidase, is sialic acid specific (*22*). To better understand the functions of the family 32 and 40 CBMs from *C. perfringens* (ATCC 13124) NanJ, we dissected the modular structure of this protein to study the structures and functions of the CBMs in isolation.

MATERIALS AND METHODS

Cloning, Protein Production and Purification. The DNA fragments encoding the family 32 and family 40 CBMs of NanJ (see Figure 1) were amplified by PCR from *C. perfringens* genomic DNA (Sigma, ATCC 13124) using previously described methods (*23*). Nucleotides 127−540 of the *nanJ* gene, corresponding to the CBM32 (amino acid residues 42−180), were amplified with the oligonucleotide primers 5′-CATATGGCTAGCGCTATTATTGAAACTGC-3′ (CBM32F) and 5′-GGATCCCTCGAGTTATTCATAAA-CATTTAACTC-3′ (CBM32R). The amplified DNA product was cloned via its 5′ and 3′ engineered *Nhe*I and *Xho*I restriction sites into identically digested pET28a to give the plasmid pCBM32. The polypeptide (called CBM32) encoded by pCBM32 comprises an H6 tag fused to the CBM32 module by a thrombin cleavage site. Nucleotides 541−1101 of the *nanJ* gene, corresponding to the CBM40 (amino acid residues 181−367), were amplified with the oligonucleotide primers 5′-CACCATCAAAGGCGAAGTAGAT-3′ (CBM40F) and 5′-CTTTTACTTAGTTTCCCCTGTTTT-3′ (CBM40R). The amplified gene fragment was ligated directly into the pET-150 TOPO Directional Cloning kit (Invitrogen, San Diego, CA) to generate pCBM40. The polypeptide (called CBM40) encoded by pCBM40 comprises an H6 tag fused to the CBM40 module by an enterokinase cleavage site.

Polypeptides were produced in 3 L cultures of *Escherichia coli* BL21(DE3) transformed with pCBM32 or pCBM40 using methods described previously (*20*) except that 50 mg/L kanamycin was used in the culture as the selective agent for pCBM32 and 100 mg/L ampicillin was used for pCBM40. CBM32 and CBM40 were purified by IMAC from cell-free extracts following previously described procedures (*20*). Purified polypeptides were concentrated and exchanged into 20 mM Tris−HCl, pH 8.0, in a stirred ultrafiltration unit (Amicon, Beverly, MA) with a 5K molecular weight cutoff (MWCO) membrane (Filtron, Northborough, MA). Purity as assessed by SDS−PAGE was greater than 95%.

Binding Studies. Carbohydrate/glycoprotein macroarrays were prepared and performed as described previously (*24*). Automated UV difference titrations were also performed essentially as described previously (*20*, *25*). Peak and trough values in the UV difference spectra at the appropriate wavelengths were extracted for further analysis. The peak-to-trough heights at two wavelength pairs were calculated by subtraction of the trough values from the peak values, and the dilution-corrected data were then plotted against the total carbohydrate concentration. Data for two wavelength pairs were analyzed simultaneously with MicroCal Origin (v.7.0) using a one-site binding model accounting for ligand depletion. Experiments were performed at 20 °C in 50 mM Tris−HCl, pH 7.5. The data reported are the averages and standard deviations of three independent titrations.

Crystallization and Data Collection. Purified samples of both CBM40 and CBM32 were exchanged into 20 mM Tris−HCl, pH 8.0. CBM40 and CBM32 were treated overnight at room temperature with enterokinase and thrombin, respectively, to remove their His_6 tags. The polypeptides were then separated from the cleaved His_6 tag by size exclusion chromatography using a Sephacryl S200 column (GE Biosciences). Pure fractions were concentrated in a 10 mL stirred ultrafiltration device using a 5K MWCO membrane. Crystallization trials were carried out using the hanging-drop vapor-diffusion method at 18 °C. Crystals of CBM40 at 20 mg/mL grew in 8% PEG 20K, 8% PEG 550 monomethyl ether, 0.2 M calcium acetate, and 0.1 M sodium acetate, pH 4.6. Crystals of CBM32 at 20 mg/mL in the presence of 20 mM galactose grew in 30% PEG 4K, 0.2 M $MgCl_2$, 0.1 M Tris−HCl, pH 8.5. Both types of protein crystals took several weeks to grow. CBM40 and CBM32 crystals were cryoprotected in mother liquor supplemented with 30% glycerol and frozen at 113 K directly in the

Table 1: Data Collection and Refinement Statistics

	CBM40	CBM32
	Data Collection	
space group	$P2_1$	$P2_12_12_1$
cell dimensions		
a, *b*, *c* (Å)	42.65, 63.11, 70.79	42.27, 45.86, 69.00
α, *β*, *γ* (deg)	90, 90.75, 90	90, 90, 90
resolution (Å)	20.00−2.20	20.00−2.25
R_{merge}	0.046 (0.257)	0.090 (0.265)
$I/\sigma(I)$	13.7 (3.7)	9.2 (4.1)
completeness (%)	92.0 (92.0)	93.6 (97.6)
redundancy	2.6 (2.4)	4.34 (4.21)
	Refinement	
resolution (Å)	20.00−2.20	20.00−2.25
no. of reflns	17720	6309
B-factor model	isotropic	isotropic
R_{work}/R_{free}	0.24/0.31	0.26/0.33
no. of atoms		
protein	1425 (chain A) 1420 (chain B)	1022
water	219	42
ligand (sugar)	21	12
B-factor		
protein	54.5 (chain A) 61.4 (chain B)	49.8
water	62.1	49.3
ligand (sugar)	54.0	48.3
rms deviation		
bond lengths (Å)	0.014	0.008
bond angles (deg)	1.676	1.253

cryostream. 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 crystal cooler. Data were processed using Crystal Clear/d*trek (*26*).

Structure Solution and Refinement. The structures of both CBM40 and CBM32 were solved by molecular replacement using the CCP4 suite of programs (*27*). Using the coordinates of the CBM40 module from the *Macrobdella decora* sialidase (PDB code 1SLI; *28*) as a search model, MOLREP (*29*) was able to find weak solutions corresponding to the two CBM40 molecules in the asymmetric unit. Initial phasing was performed by five cycles of rigid body refinement with REFMAC (*30*) followed by solvent flatting and NCS averaging with DM using a conservative protocol of phase extension from 4 Å over 500 cycles (*31*). The initial model was then manually altered using COOT (*32*) to fit the electron density. Successive rounds of refinement with REFMAC and model building with COOT were required to complete the model.

The structure solution of CBM32 was straightforward using MOLREP to find the single molecule in the asymmetric unit using the CBM32 from *C. perfringens* GH84C as a search model (PDB code 2J1A; *20*). Successive rounds of refinement with REFMAC and model building with COOT were used to manually correct the model.

In both cases, water molecules were added using the ARP/wARP (*33*) option within REFMAC and inspected visually prior to deposition. For each data set 5% of the observations were flagged as "free" (*34*) and used to monitor the refinement procedures. All data collection, processing, and final model statistics are given in Table 1. The structure factors and coordinates for CBM32 and CBM40 have been deposited with PDB codes of 2v72 and 2v73, respectively.

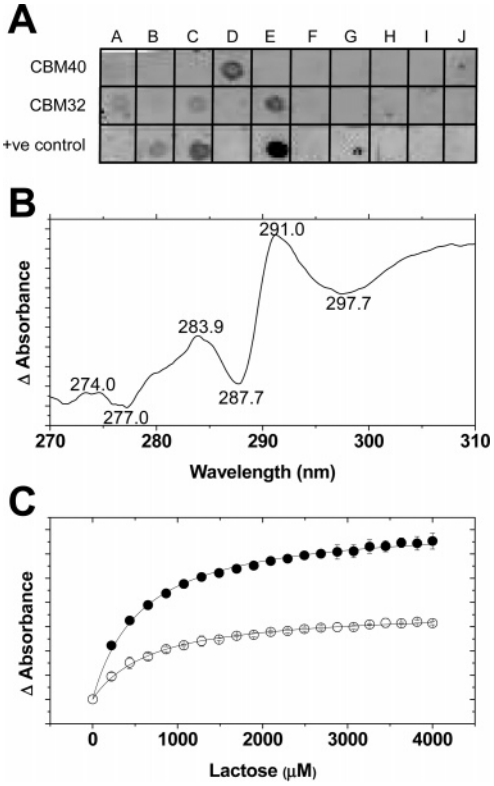

FIGURE 2: Binding properties of CBM32 and CBM40 from *C. perfringens* NanJ. (A) Glycoprotein/polysaccharide macroarray binding analysis of CBM32 and CBM40. The positive control was the previously characterized family 32 CBM from the *C. perfringens* GH84C (*20*). The glycoproteins and polysaccharides tested were (A) mucin type I-S, (B) mucin type II, (C) mucin type III, (D) fetuin, (E) asialofetuin, (F) thyroglobulin, (G) hyaluronic acid, (H) chondroitin sulfate, (I) heparin, and (J) proteoglycan. (B) representative UV difference spectrum taken from the final addition of lactose to 30 μM CBM32 in a quantitative UV difference titration. Wavelengths at which peaks or troughs occur are labeled. (C) Representative UV difference binding isotherm for D-galactose titrated into 30 μM CBM32. Closed symbols show the isotherm produced by using the 291.0−287.7 nm wavelength pair. Open symbols show the isotherm produced by using the 291.0−297.7 nm wavelength pair. Error bars represent the standard deviations of data obtained from three independent titrations. The solid line shows the fits to a one-site binding model assuming a perfect 1:1 stoichiometry.

RESULTS AND DISCUSSION

CBM32 and CBM40 from NanJ Display Differing Binding Specificities. The binding abilities of CBM32 and CBM40 were tested by macroarray screening against a small panel of glycoproteins and polysaccharides. CBM40 bound only to fetuin, while CBM32 bound to mucin I−S, mucin III, and asialo fetuin (Figure 2A). The binding of CBM40 to fetuin but not asialofetuin, which only lacks the terminal sialic acid residues, is consistent with a specificity for sialic acid. Quantitation of the interactions between CBM40 and sialic acid and fetuin was pursued by isothermal titration calorimetry. Though heats were evolved in these experiments, indicating an interaction (not shown), the binding was too weak to accurately quantify by this method. The was in contrast to the family 40 sialic acid specific CBM from the *Vibrio* sialidase which bound sialic acid with a dissociation constant in the micromolar range (*22*). The specificity of CBM32 for the tested glycoproteins was similar to that of the *N*-acetyllactosamine (lacNAc) specific CBM32 from *C.*

perfringens GH84C, which was used as the positive control, suggesting that it also requires terminal galacto-configured sugar residues. The interaction of CBM32 with small soluble sugars was investigated further by UV difference absorption spectroscopy. Out of a panel of sugars comprising mannose, fucose, galactose, glucose, sialic acid, *N*-acetylglucosamine (glcNAc), *N*-acetylgalactosamine (galNAc), and lactose, only galactose, galNAc, and lactose gave UV difference signals (not shown). The positions of the peaks and troughs in the UV difference spectra indicated the involvement of tryptophan in the recognition of galactose (Figure 2B) (*35*), which agreed with the crystal structure (see below). This was followed up quantitatively by UV difference titrations (Figure 2C) which revealed association constants of 1.36 ($\pm$0.13) $\times$ 10^3, 4.38 ($\pm$0.21) $\times$ 10^3, and 1.77 ($\pm$0.18) $\times$ 10^3 M^{-1} for galactose, galNAc, and lactose, respectively.

Overall, the results indicate a preference of CBM32 for terminal galacto-configured sugars and a preference of CBM40 for sialic acid. These ligand preferences are entirely consistent considering that CBM32 and CBM40 belong to families that are generally galactose and sialic acid specific, respectively (*18*−*22*). The affinities of these CBMs for monosaccharides appear to be quite low in the range of 10^3 M^{-1} for CBM32 and likely in a similar range for CBM40 binding to sialic acid. Attempts to find higher affinity ligands by glycan microarray screening through Core H at the Consortium for Functional Glycomics were unsuccessful, likely due to low binding affinities.

The Structure of CBM32 from NanJ Reveals a Simple Galactose-Binding Site. Crystals of CBM32 complexed with galactose grew in the space group $P2_12_12_1$ over a period of roughly 6 months. The structure of this protein was solved by molecular replacement and refined at a resolution of 2.25 Å (see Table 1 for model statistics). This module adopts a β-sandwich fold comprising a three-stranded antiparallel β-sheet opposing a five-stranded antiparallel β-sheet (Figure 3A). CBM32 coordinates one metal ion, which was modeled as Ca^{2+}, at a binding site that is highly conserved among family 32 CBMs and several other families of CBMs (*18*, *20*) (Figure 3A). This metal is presumed to be a structural component and not involved in carbohydrate binding due to its distance from the binding site (Figure 3A). The oxygen atoms coordinating the metal are contributed from the side chains of D71, T76, and E175 and the backbone carbonyls of S68, N73, T76, and A174 (not shown).

Clear electron density for a single molecule of β-D-galactose was present in the expected binding site of CBM32, allowing it to be suitably modeled (Figure 3B). This revealed remarkably sparse interactions between the CBM and the galactose. Direct hydrogen bonds appeared to be limited to two hydrogen bonds between the O3 hydroxyl group of galactose and the terminal nitrogens of R110 and a single hydrogen bond between the ϵN of H79 and the O4 hydroxyl group (Figure 3C). A single water-mediated hydrogen bond is made between D64 and the O3 hydroxyl group. A primary interaction is the packing of the planar indole rings of W82 against the flat, apolar surface created by carbons 3−6 on the B-face of D-galactose, allowing the O4 to point away from the aromatic residue (Figure 3B). F170 is situated in the binding site at roughly right angles to W82, which creates an apolar pocket that accommodates the C6 group of galactose (Figure 3B).

Overall, the binding site of CBM32 is a shallow slot that is blocked off at the end where the C6 hydroxyl methyl group of galactose sits (Figure 3D). The hydrogen-bonding pattern suggests a strict specificity for an axial O4 and an equatorial O3. In contrast, the hydroxyl groups on the C1 and C2 of the bound galactose are very solvent exposed. This is consistent with binding to asialofetuin, which has an abundance of terminal galactoses linked by glycosidic bonds at C1. Furthermore, due to the shallow nature of the binding site, any substituent linked to the C1 of galactose would protrude away from the protein, suggesting a lack of subsites on the protein that recognize sugar residues preceding the galactose. The exposure of the O2 group in the galactose bound to CBM32 suggests that it may accommodate modifications at C2. Indeed, this is consistent with its capacity to bind galNac, which has a common acetamido modification at C2. However, in the absence of a CBM32−galNac complex we are unable to determine whether the acetamido group would contribute any additional interactions to explain the slight preference of CBM32 for galNac. Thus, it is unclear whether the measured preference for galNac is real and biologically relevant. Like the CBM32 from GH84C of *C. perfringens*, CBM32 may also accommodate the H-blood group antigen determinant, which is a fucosyl residue linked α(1−2) to galactose. However, it is currently unknown whether this is true.

Galactose residues are often modified with α(2−3)- or α-(2−6)-linked sialic acid residues, such as in fetuin (*36*). In the CBM32−galactose complex the C3 hydroxyl group is buried in the binding site and is responsible for the majority of the protein−carbohydrate hydrogen bonds (Figure 3B−D). Thus, CBM32 is unlikely to tolerate α(2−3)-linked sialic acids. The C6 hydroxyl group does not make any hydrogen bonds; however, it is tucked into a pocket in the binding site with little room allowed for modification by an α(2−6)-linked sialic acid residue (Figure 3D). This is consistent with CBM32's inability to bind fetuin.

Overall, the structural data suggest that CBM32 requires a terminal galactose (or possibly galNac), with the possibility of accommodating the H-antigen, but may have considerable flexibility in the sugar residues that precede the galactose in the receptor glycan. This agrees with the lack of strong hits from microarray screening, where the low affinity of binding would result in a number of weak hits that are difficult to discern from the background. It appears that this CBM has a very simple galactose recognition site, the architecture of which is very well conserved with the CBM32's from the *Cladobotryum dendroides* galactose oxidase (Figure 3E) and the *Micromonospora viridifaciens* sialidase (not shown). These CBMs may represent the prototypic galactose binding CBMs within the family. In contrast, other CBM32's, for example, the CBM32 from *C. perfringens* GH84C, have evolved more elaborate binding site architectures that create additional subsites and greater specificity for complex glycans while maintaining the core amino acid side chains that interact with the key galactose residue (Figure 3F) (*20*).

Structure of CBM40 from NanJ. CBM40 was very recalcitrant to crystallization. Ultimately, a single protein preparation yielded one crystal of reasonable quality. This crystal, which was of the space group $P2_1$ with two molecules in the asymmetric unit, gave a good-quality data set to 2.2

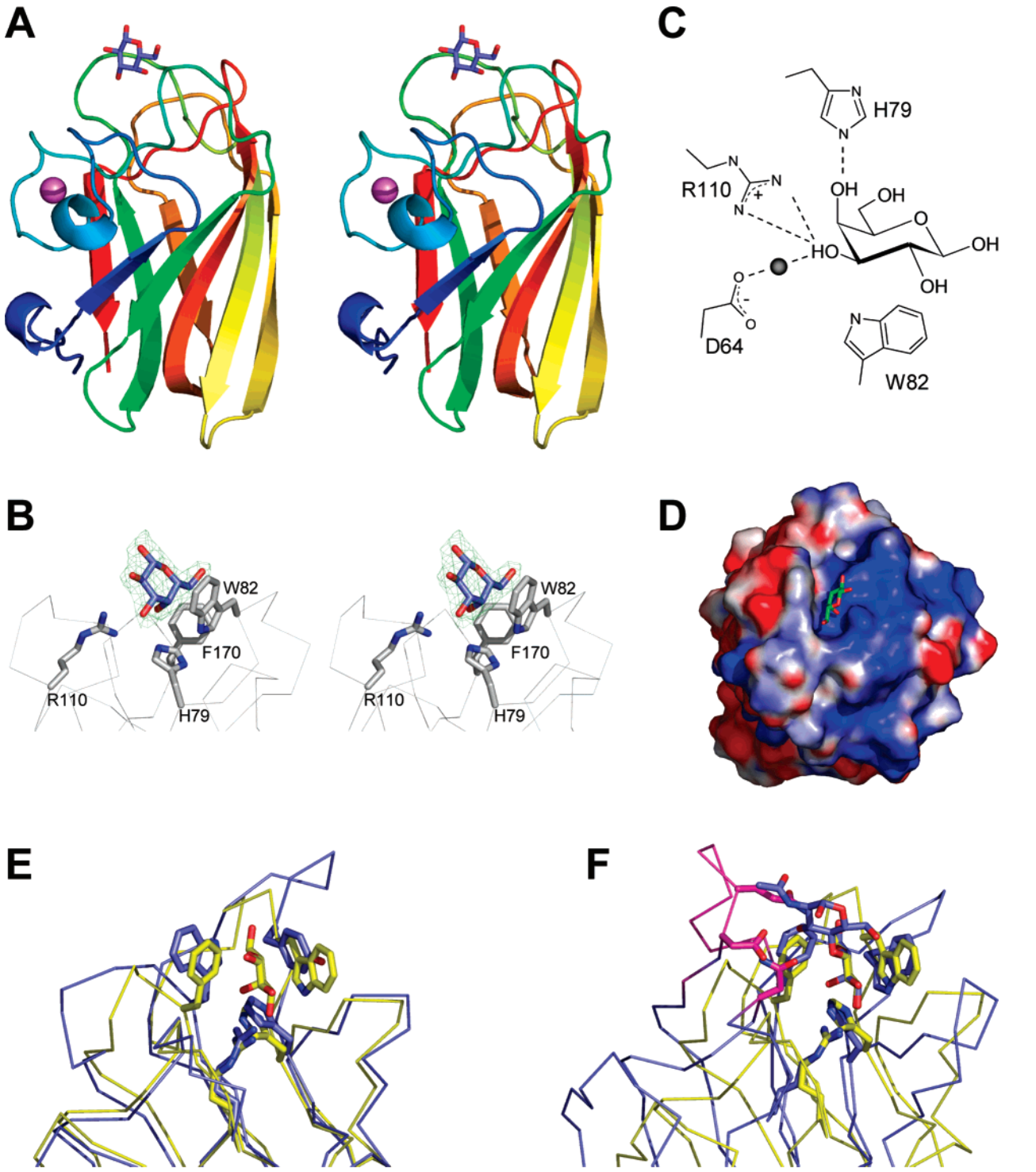

FIGURE 3: Structure of CBM32 from *C. perfringens* NanJ. (A) Stereoview of the "cartoon" representation of CBM32 with the bound calcium ion shown as a violet sphere and the bound galactose shown in stick representation. (B) Galactose binding site of CBM32 with the galactose (blue) and the side chains (gray) that interact with the sugar shown in stick representation. The galactose is surrounded by a maximum-likelihood (*30*)/σ_A (*42*) weighted $2F_o - F_c$ electron density map contoured at 1σ (0.1 electron/Å^3). (C) Schematic showing the interaction of CBM32 with galactose. A distance of 3.2 Å was used as the cutoff for the assignment of putative hydrogen bonds. A water molecule is shown as a shaded sphere. Protons of the amino acid side chains were omitted for clarity. (D) Solvent-accessible surface area of CBM32 colored by relative electrostatic potential showing the galactose (green stick representation) in the binding site. (E) Overlay of CBM32 (yellow) with the CBM32 from the *C. dendroides* galactose oxidase (blue). Conserved binding site residues and ligands are shown in stick representation. (F) Overlay of CBM32 (yellow) with the CBM32 from the *C. perfringens* GH84C (blue and pink). Because the binding sites are partially offset in these two proteins, the overlay was produced by overlapping the analogous galactose residues in each of the binding sites to better visualize the conserved binding residues. Conserved binding site residues and ligands are shown in stick representation. The extended loop in the CBM32 from the *C. perfringens* GH84C that contributes to its specificity for *N*-acetyllactosamine is shown in pink.

Å. The structure of CBM40 was then solved by molecular replacement (see Table 1 for model statistics).

CBM40 adopts a β-sandwich fold comprising a six-stranded antiparallel β-sheet opposing a five-stranded anti-

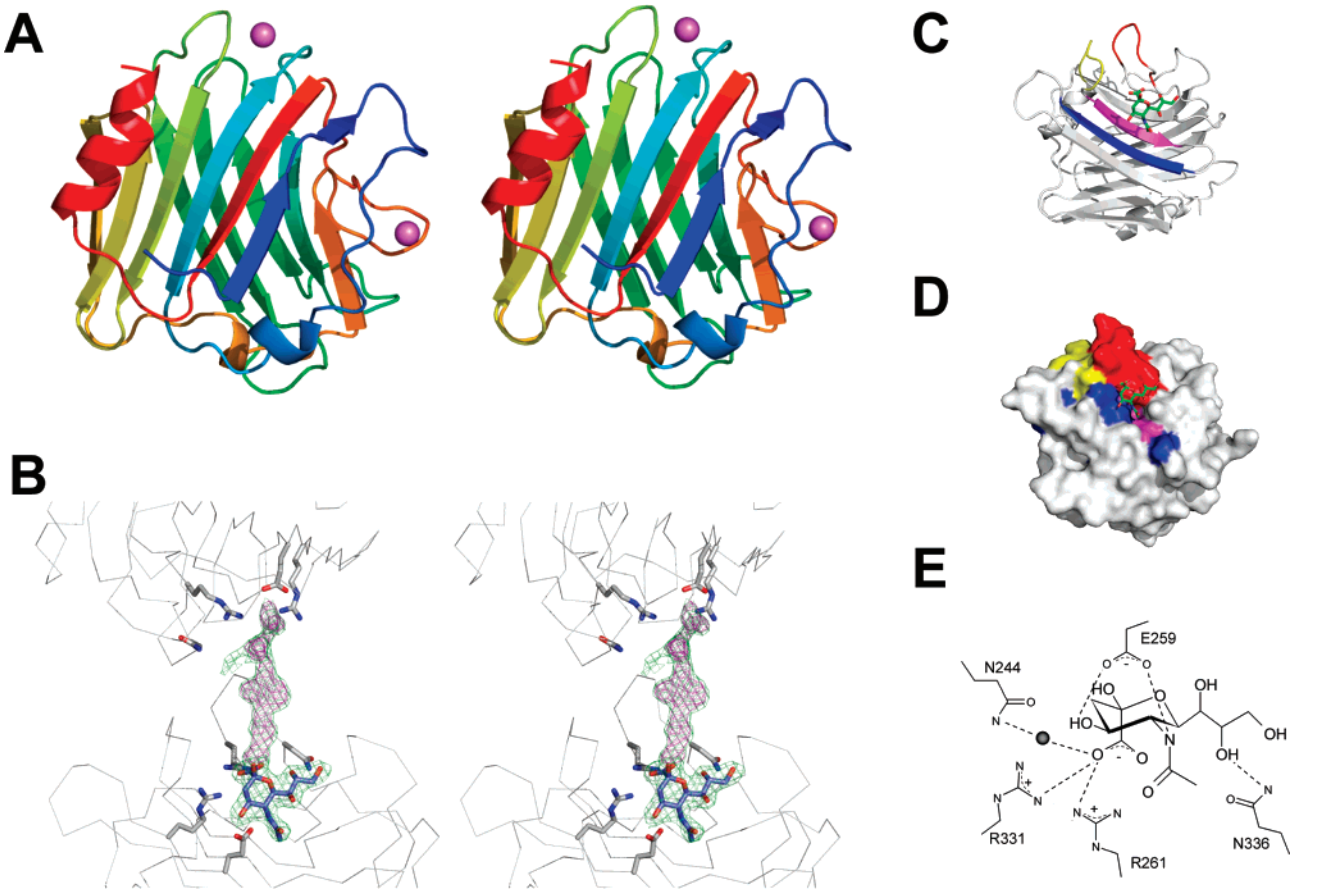

FIGURE 4: Structure of CBM40 from *C. perfringens* NanJ. (A) Stereoview of the "cartoon" representation of CBM40 with the bound calcium ions shown as violet spheres. (B) Stereoview of the ligand-mediated dimerization of CBM40. The bound sialic acid (blue) and side chains (gray) that bind it are shown in stick representation. The electron density for the full ligand is shown as a green maximum-likelihood (*30*)/σ_A (*42*) weighted $2F_o - F_c$ electron density map contoured at 1σ (0.1 electron/Å^3). The difference density for the unmodeled portion of the ligand is shown as a magenta maximum-likelihood/σ_A (*42*) weighted $F_o - F_c$ map contoured at 3σ (0.06 electron/Å^3). (C). A cartoon representation showing the secondary structure elements that create the sialic acid binding site. The loop separating β-strands 4 and 5 is shown in yellow, the loop separating β-strands strands 11 and 12 is shown in red, β-strand 5 is shown in pink, and β-strand 6 is shown in blue. (D) Solvent-accessible surface area of CBM40. The surface color corresponds to that in panel C. (E) schematic showing the interaction of CBM40 with sialic acid. A distance of 3.2 Å was used as the cutoff for the assignment of putative hydrogen bonds. A water molecule is shown as a shaded sphere. Protons of the amino acid side chains were omitted for clarity.

parallel β-sheet with a small C-terminal α-helix packing against the six-stranded sheet (Figure 4A). Two metal atoms were found bound to molecule A (Figure 4A). These were judged most likely to be Ca^{2+} on the basis of the high concentration of $CaCl_2$ in the crystallization conditions but unlikely to be biologically relevant as the atoms were not found bound to molecule B. DALI (*37*) searches with the CBM40 coordinates produced the best hits with domains from the *M. decora* sialidase (PDB code 2SLI; *28*), the *V. cholerae* sialidase (PDB code 1W0P; *22*), and a *Trypanosoma rangeli* sialidase (PDB code 1MZ5; *38*) (see below).

As the refinement and building progressed to the point where both molecules in the asymmetric unit were completely built, an unexpected region of strong electron density not corresponding to any portions of the protein was found (Figure 4B). Interestingly, molecules A and B in the asymmetric unit were being bridged by this electron density. Furthermore, the ends of the molecule represented by this electron density appeared to be bound at the same sites on molecules A and B of CBM40 (Figure 4B). The end of this electron density that terminated at molecule A was very clearly a sialic acid, and it was facile to model it as such (Figure 4B). The bulge of density in the middle appeared to resemble an additional pyranose sugar linked to the terminal sialic acid via a 3−4 atom linker. This led us to believe that the CBM40 had fortuitously cocrystallized with a fragment of the *E. coli* polysialic acid capsule, which in the BL21 strain appears to have alternating α(2−8) and α(2−9) bonds (*39*). Unfortunately, though the terminus at molecule A is clearly a sialic acid, the shape of the electron density neighboring was inconsistent with sialic acid, suggesting the ligand was not α(2−8)- or α(2−9)-linked sialic acid. Thus, we were unable to identify the complete molecule with which CBM40 fortuitously cocrystallized; however, it clearly appears to terminate with a sialic acid that is linked to an unidentified pyranose sugar. This pyranose sugar is itself attached to another unidentified residue that interacts with the binding site in molecule B. Despite our inability to identify the complete molecule, this happenstance was of exceptionally good fortune, not only because the molecule appears to be critical in forming the crystal lattice but because it is consistent with CBM40's apparent specificity for sialylated glycans and provides some insight into the molecular details of ligand recognition.

Divergent Modes of Sialic Acid Recognition in CBM40's. The sialic acid binding site of CBM40 is created mainly from

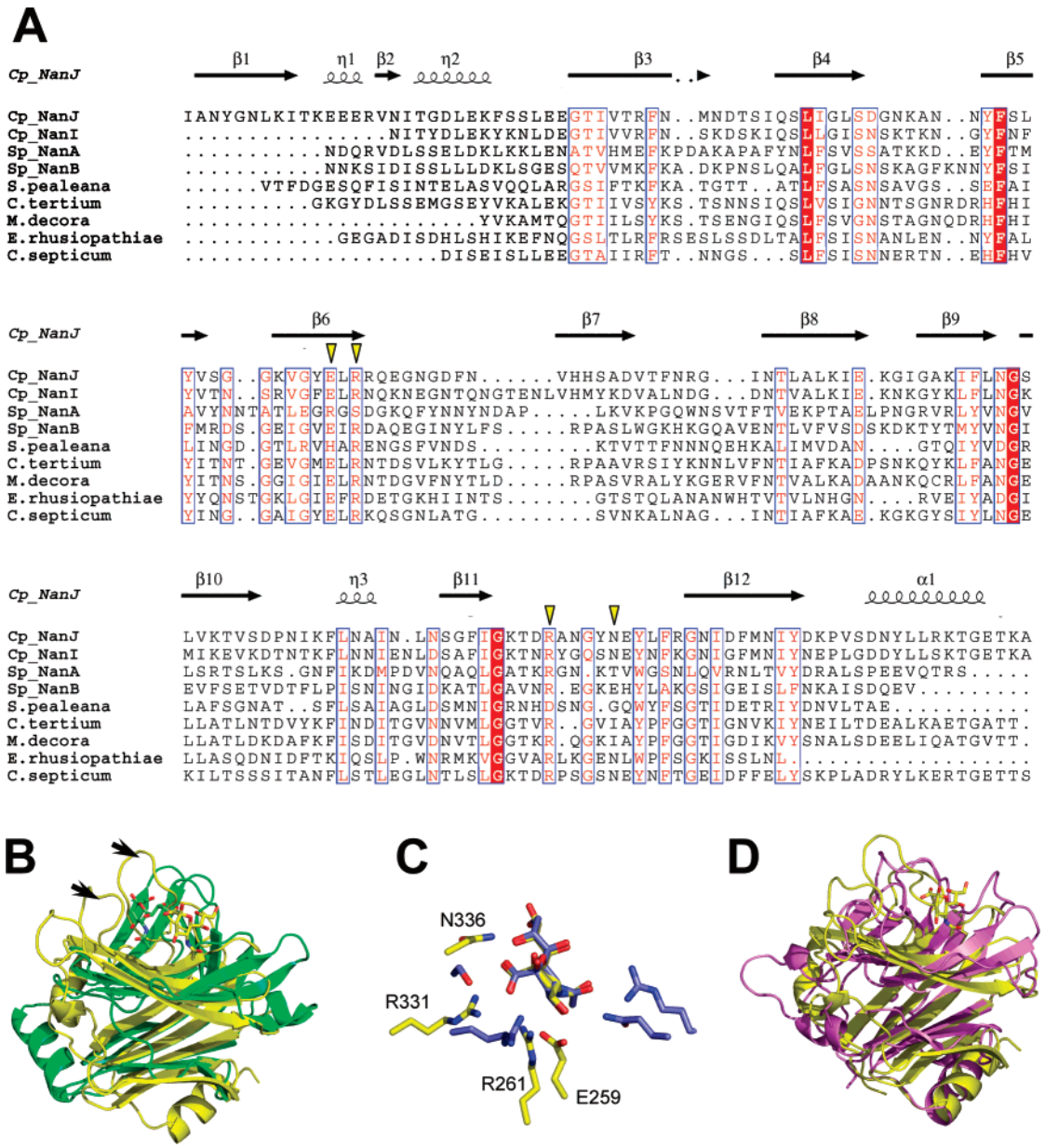

FIGURE 5: Comparison of CBM40 with structurally related polypeptides. (A) Sequence alignment of CBM40 from NanJ (shown as Cp_NanJ) with related modules from NanI (Cp_NanI), *Streptococcus pneumoniae* NanA and NanB (Sp_NanA and Sp_NanB, respectively), and sialidases from *S. pealeana*, *Clostridium tertium*, *M. decora*, *Erysipelothrix rhusiopathiae*, and *Clostridium septicum*. The secondary structure elements of CBM40 are indicated above the alignment. Relevant residues in CBM40 that are involved in binding sialic acid are indicated with inverted yellow triangles. This panel was generated with ESPript (*43*). (B) Overlay of CBM40 (yellow) with the CBM40 from the *V. cholerae* sialidase (green). Bound sialic acid residues are shown as stick models. The loops in CBM40 that form a portion of the binding site discussed in the text are indicated with arrows. (C) Structural overlap of the sialic acid binding sites from CBM40 (yellow) and the CBM40 from the *V. cholerae* sialidase (blue). The overlay was produced by overlapping equivalent atoms in the bound sialic acid residues. Relevant amino acids involved in coordinating the sugars are shown in stick representation. (D) Structural overlap of CBM40 (yellow) with the CBM40-like module from the *T. rangeli trans*-sialidase (purple).

the surfaces contributed by two loops and β-strands 5 and 6 (Figure 4C,D). The loops separating β-strands 4 and 5 (residues 239−255) and β-strands 11 and 12 (residues 327−342) contact one another and extend away from the protein to make a surface that is roughly perpendicular to the β-sheets of the main fold. This surface is relatively flat, allowing the planar surface of the sialic acid pyranose ring to sit up against it (Figure 4D). The specificity of CBM40 for this sugar is driven by the specific protein−carbohydrate interactions that are schematically represented in Figure 4E. Briefly, β-strand 6 contains E259, which forms direct hydrogen bonds with the O4 and N5 (acetamido nitrogen) of sialic acid. Also in β-strand 6 is R261, which may form a salt bridge with the axial C1 carboxylate group of the sialic acid. The loop comprising residues 239−255 harbors N244, whose side chain is involved in a water-mediated hydrogen bond with the C1 carboxylate group. The loop comprising residues 327−342 contains R331, which may also salt bridge to the C1 carboxylate group, and N336, whose side chain makes a hydrogen bond with O8. Though we have suggested the possible formation of protein−ligand ion−ion interactions on the basis of the structure and ionization states at the experimental pH, the apparently low affinity of the interaction does not bear out this type of interaction. Given the

relative strength of ionic interactions, the binding of sialic acid to CBM40 would be expected to be stronger than that measured, suggesting that it is not an ion pair interaction but simply hydrogen bonding.

The most closely related family 40 CBM is the module from the *M. decora* sialidase (PDB code 1SLI; *28*). This CBM shares ~28% amino acid sequence identity with CBM40 and overlaps with an rmsd of 1.5 Å over 173 matched Cα's (measured with the SSM algorithm (*40*) as implemented in COOT (*32*)) (not shown). Residues in the binding site implicated in direct protein−carbohydrate interactions are highly conserved among the *M. decora* CBM, CBM40, and several other CBM40's originating from a variety of other organisms, suggesting that they all share a similar mode of sialic acid recognition (Figure 5A). More distantly related is the N-terminal module of the *V. cholerae* sialidase, which is also classified into CBM family 40 but shares only ~11% amino acid identity with CBM40. The *Vibrio* CBM40 is the only module classified into family 40 that has previously been demonstrated to bind sialic acid (*22*). This module overlaps CBM40 with an rmsd of 3.2 Å over 123 matched Cα's (measured with the SSM algorithm (*40*) as implemented in COOT (*32*)). A structural comparison of these CBMs shows the location of the binding sites to be somewhat different (Figure 5B). The loops that form a portion of the binding site in CBM40 are missing from the *Vibrio* CBM40 such that the binding site on this CBM is a shallow groove formed between the two shortened loops (Figure 5B). An overlap of the binding sites further highlights the differences between these two sialic acid binding CBM40's, which revealed there is little to no structural conservation in the active site residues of these CBMs (Figure 5C). Within CBM family 40, there are two other modules, both from the *Shewanella pealeana* sialidase, that show reasonably close sequence identity with the *Vibrio* CBM40 and that have all of the residues involved in sialic acid binding conserved with the *Vibrio* CBM40 (not shown). Thus, there appear to be two subfamilies in CBM family 40—one typified by the *Clostridium* CBM40 and the other by the *Vibrio* CBM40—that have binding sites that are in different locations and lack conserved binding residues. This is quite unusual for CBMs, which usually have a conserved binding site location throughout a family (the family 6 CBM from *Cellvibrio mixtus* might be considered an exception as it has a second binding site in addition to the one that is conserved among family members (*41*)).

CBM40 also showed structural similarity to the accessory modules in trypanosomal sialidases. The C-terminal module of the *T. rangeli* sialidase overlaps CBM40 with an rmsd of 3.0 Å over 153 matched Cα's (measured with the SSM algorithm (*40*) as implemented in COOT (*32*)) (Figure 5D) and an amino acid sequence identity of ~9%. None of the sialic acid binding residues are conserved between this module and CBM40 or the *Vibrio* CBM40. Furthermore, it is currently unknown whether the trypanosomal modules bind to carbohydrates; however, its context and structural relatedness to known sialic acid binding CBMs suggest that the trypanosomal modules may constitute a distinct family of sialic acid binding modules.

Multimodularity in the Large Clostridial Sialidases. Strains of *C. perfringens* produce three sialidases ranging in modularity from one module up to six modules. The single-module NanH is found intracellularly and is the only sialidase produced in the gastroenteritis-causing SM101 strain. NanI and NanJ have two and six modules, respectively, and are secreted only by the myonecrotic strain 13 and ATCC 13124. Here we have shown that the CBM40 in NanJ is indeed a sialic acid binding CBM. Given the degree of sequence identity between CBM40 from NanJ and CBM40 from NanI (56% amino acid sequence identity) and the conservation of binding site residues (Figure 5A), it is likely that the CBM from NanI also binds sialic acid. Relative to NanI, NanJ possesses the additional N-terminal CBM32 that binds terminal galacto-configured sugars (i.e., galactose and galNAc). Both CBM32 and CBM40 appear to have very low affinities, and we tentatively suggest on the basis of the structures and glycan screening evidence that their specificities are not significantly influenced by sugars other than the terminal sugars of the glycan. Thus, it seems plausible that CBM32 and CBM40 work in tandem to bind avidly to glycoproteins or cell surfaces that display populations of clustered glycans having both terminal galactoses/galNAc's and sialic acids. Even though the independent CBMs have low affinity, the simultaneous association of the tethered modules in the parent enzyme, NanJ, could result in reasonably tight binding as a result of this multivalent interaction. Precisely why an enzyme with specificity for sialic acid would require nonsialylated glycans as part of its receptor is unclear. This may relate to the functions of the three C-terminal modules of unidentified function. The last two of these modules, X82 (cohesin-like) and FN3 (fibronectin type 3-like), show distant amino acid sequence identity with modules involved in protein−protein interactions. It is possible that these modules are involved in recruiting other clostridial enzymes of differing specificities to the carbohydrate-bearing surface, thus requiring that NanJ remain, albeit very weakly, adhered to the surface even after it has been mostly desialylated. In contrast, NanI, which only has the sialic acid specific CBM, does not possess the additional modules that may contribute to enzyme recruitment, suggesting a role devoted only to desialylation. Overall, the CBMs in NanI and NanJ would increase the adherence of the enzymes to glycan-bearing surfaces, thus promoting efficient hydrolysis of their target sugars. In turn, this would enhance virulence by aiding in tissue destruction by other virulence factors (other hydrolases and the α-toxin).

ACKNOWLEDGMENT

We thank Alicia Lammerts van Bueren for technical assistance. We also thank Core H of the Consortium of Functional Glycomics for their efforts with the glycan array screening.

REFERENCES

1. Petit, L., Gibert, M., and Popoff, M. R. (1999) *Clostridium perfringens*: toxinotype and genotype, *Trends Microbiol. 7*, 104−10.
2. Rood, J. I. (1998) Virulence genes of *Clostridium perfringens*, *Annu. Rev. Microbiol. 52*, 333−60.
3. Van Immerseel, F., De Buck, J., Pasmans, F., Huyghebaert, G., Haesebrouck, F., and Ducatelle, R. (2004) *Clostridium perfringens* in poultry: an emerging threat for animal and public health, *Avian Pathol. 33*, 537−49.
4. Rood, J. I., and Wilkinson, R. G. (1976) Relationship between hemagglutinin and sialidase from *Clostridium perfringens*

CN3870: chromatographic characterization of the biologically active proteins, *J. Bacteriol. 126*, 831−44.
5. Rood, J. I., and Wilkinson, R. G. (1976) Relationship between hemagglutinin and sialidase from *Clostridium perfringens* CN3870: gel filtration of mutant and reverant activities, *J. Bacteriol. 126*, 845−51.
6. Roggentin, P., Gutschker-Gdaniec, G. H., Hobrecht, R., and Schauer, R. (1988) Early diagnosis of clostridial gas gangrene using sialidase antibodies, *Clin. Chim. Acta 173*, 251−62.
7. Roggentin, P., Rothe, B., Lottspeich, F., and Schauer, R. (1988) Cloning and sequencing of a *Clostridium perfringens* sialidase gene, *FEBS Lett. 238*, 31−4.
8. Traving, C., Schauer, R., and Roggentin, P. (1994) Gene structure of the 'large' sialidase isoenzyme from *Clostridium perfringens* A99 and its relationship with other clostridial nanH proteins, *Glycoconjugate J. 11*, 141−51.
9. Flores-Diaz, M., Alape-Giron, A., Clark, G., Catimel, B., Hirabayashi, Y., Nice, E., Gutierrez, J. M., Titball, R., and Thelestam, M. (2005) A cellular deficiency of gangliosides causes hypersensitivity to *Clostridium perfringens* phospholipase C, *J. Biol. Chem. 280*, 26680−9.
10. Klein, R. L., Novak, R. W., and Novak, P. E. (1986) T-cryptantigen exposure in neonatal necrotizing enterocolitis, *J. Pediatr. Surg. 21*, 1155−8.
11. Placzek, M. M., and Gorst, D. W. (1987) T activation haemolysis and death after blood transfusion, *Arch. Dis. Child. 62*, 743−4.
12. Batge, B., Filejski, W., Kurowski, V., Kluter, H., and Djonlagic, H. (1992) Clostridial sepsis with massive intravascular hemolysis: rapid diagnosis and successful treatment, *Intensive Care Med. 18*, 488−90.
13. Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., and Hayashi, H. (2002) Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater, *Proc. Natl. Acad. Sci. U.S.A. 99*, 996−1001.
14. Myers, G. S., Rasko, D. A., Cheung, J. K., Ravel, J., Seshadri, R., DeBoy, R. T., Ren, Q., Varga, J., Awad, M. M., Brinkac, L. M., Daugherty, S. C., Haft, D. H., Dodson, R. J., Madupu, R., Nelson, W. C., Rosovitz, M. J., Sullivan, S. A., Khouri, H., Dimitrov, G. I., Watkins, K. L., Mulligan, S., Benton, J., Radune, D., Fisher, D. J., Atkins, H. S., Hiscox, T., Jost, B. H., Billington, S. J., Songer, J. G., McClane, B. A., Titball, R. W., Rood, J. I., Melville, S. B., and Paulsen, I. T. (2006) Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*, *Genome Res. 16*, 1031−40.
15. Henrissat, B., and Davies, G. J. (2000) Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics, *Plant Physiol. 124*, 1515−1519.
16. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine tuning polysaccharide recognition, *Biochem. J. 382*, 769−782.
17. Thobhani, S., Ember, B., Siriwardena, A., and Boons, G. J. (2003) Multivalency and the mode of action of bacterial sialidases, *J. Am. Chem. Soc. 125*, 7154−5.
18. Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R., and Davies, G. (2003) Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and "galactose-binding" domains, *J. Mol. Biol. 327*, 659−69.
19. Gaskell, A., Crennell, S., and Taylor, G. (1995) The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll, *Structure 3*, 1197−1205.
20. Ficko-Blean, E., and Boraston, A. B. (2006) The interaction of carbohydrate-binding module from a *Clostridium perfringens* N-acetyl-beta-hexosaminidase with its carbohydrate receptor, *J. Biol. Chem. 281*, 37748−57.
21. Newstead, S. L., Watson, J. N., Bennet, A. J., and Taylor, G. (2005) Galactose recognition by the carbohydrate-binding module of a bacterial sialidase, *Acta Crystallogr., D: Biol. Crystallogr. 61*, 1483−91.
22. Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., Kiefel, M. J., von Itzstein, M., and Taylor, G. (2004) Sialic acid recognition by *Vibrio cholerae* neuraminidase, *J. Biol. Chem. 279*, 40819−26.
23. Ficko-Blean, E., and Boraston, A. B. (2005) Cloning, recombinant production, crystallization and preliminary X-ray diffraction studies of a family 84 glycoside hydrolase from *Clostridium perfringens*, *Acta Crystallogr., F: Struct. Biol. Cryst. Commun. 61*, 834−6.
24. Lammerts van Bueren, A., Higgins, M., Wang, D., Burke, R. D., and Boraston, A. B. (2007) Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors, *Nat. Struct. Mol. Biol. 14*, 76−84.
25. Boraston, A. B., Warren, R. A., and Kilburn, D. G. (2001) beta-1,3-Glucan binding by a thermostable carbohydrate-binding module from *Thermotoga maritima*, *Biochemistry 40*, 14679−14685.
26. Pflugrath, J. W. (1999) The finer things in X-ray diffraction data collection, *Acta Crystallogr., D: Biol. Crystallogr. 55*, Part 10, 1718−25.
27. Collaborative Computational Project Number 4. (1994) The CCP4 suite−programs for protein crystallography, *Acta Crystallogr., D: Biol. Crystallogr. 50*, 760−763.
28. Luo, Y., Li, S. C., Chou, M. Y., Li, Y. T., and Luo, M. (1998) The crystal structure of an intramolecular trans-sialidase with a NeuAc alpha2 → 3Gal specificity, *Structure 6*, 521−30.
29. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr., D: Biol. Crystallogr. 56*, Part 12, 1622−4.
30. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum likelihood method, *Acta Crystallogr., D: Biol. Crystallogr. 53*, 240−255.
31. Cowtan, K. D., and Zhang, K. Y. (1999) Density modification for macromolecular phase improvement, *Prog. Biophys. Mol. Biol. 72*, 245−270.
32. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr., D: Biol. Crystallogr. 60*, 2126−32.
33. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
34. Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472−475.
35. Boraston, A. B., Chiu, P., Warren, R. A. J., and Kilburn, D. G. (2000) Specificity and Affinity of Substrate Binding by a Family 17 Carbohydrate-Binding Module from *Clostridium cellulovorans* Cellulase 5A, *Biochemistry 39*, 11129−11136.
36. Takasaki, S., and Kobata, A. (1986) Asparagine-linked sugar chains of fetuin: occurrence of tetrasialyl triantennary sugar chains containing the Gal beta 1−3GlcNAc sequence, *Biochemistry 25*, 5709−15.
37. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123−138.
38. Buschiazzo, A., Tavares, G. A., Campetella, O., Spinelli, S., Cremona, M. L., Paris, G., Amaya, M. F., Frasch, A. C., and Alzari, P. M. (2000) Structural basis of sialyltransferase activity in trypanosomal sialidases, *EMBO J. 19*, 16−24.
39. Andreishcheva, E. N., and Vann, W. F. (2006) Escherichia coli BL21(DE3) chromosome contains a group II capsular gene cluster, *Gene 384*, 113−9.
40. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr., D: Biol. Crystallogr. 60*, 2256−68.
41. Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M., Ferreira, L. M., Bornet, O., Zamboni, V., Darbon, H., Smith, N. L., Black, G. W., Henrissat, B., and Gilbert, H. J. (2001) The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved, *J. Biol. Chem. 276*, 48580−48587.
42. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors., *Acta Crystallogr., A: Found. Crystallogr. 42*, 140−149.
43. Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins, *Nucleic Acids Res. 31*, 3320−3.

BI701317G