Portrait of an Enzyme, a Complete Structural Analysis of a Multimodular $\beta$-N-Acetylglucosaminidase from 
Clostridium perfringens$^{*5}$

Elizabeth Ficko-Blean$^{1}$, Katie J. Gregg$^{4}$, Jarrett J. Adams$^{1,2}$, Jan-Hendrik Hehemann$^{6}$, Mirjam Czjzek$^{6}$, Steven P. Smith$^{3,7}$, and Alisdair B. Boraston$^{1,4}$

From the $^{1}$Department of Biochemistry & Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada, the $^{2}$Department of Biochemistry, Queen’s University, Kingston, Ontario K7L 3N6, Canada, and the $^{3}$Station Biologique de Roscoff, Marine Plants and Biomolecules, CNRS-University Pierre and Marie Curie Paris 6, BP74, Roscoff 29682, France

Common features of the extracellular carbohydrate-active virulence factors involved in host-pathogen interactions are their large sizes and modular complexities. This has made them recalcitrant to structural analysis, and therefore our understanding of the significance of modularity in these important proteins is lagging. Clostridium perfringens is a prevalent human pathogen that harbors a wide array of large, extracellular carbohydrate-active enzymes and is an excellent and relevant model system to approach this problem. Here we describe the carbohydrate-active enzymes and is an excellent and relevant model system to approach this problem. Here we describe the carbohydrate-active enzymes of the full-length enzymes, knowledge that would have considerable insight into the individual functions of the modules.

Carbohydrate-active enzymes involved in processing eukaryotic glycans is their structural complexity; such proteins are frequently very large (over 1000 amino acids) and can comprise numerous modules and domains (12–16). C. perfringens, a myonecrotic ("flesh-eating") bacterial pathogen and a leading cause of food-borne gastrointestinal illness, deploys numerous protein exo-toxins that work together to increase the pathogen’s virulence (3, 17). Among its exo-toxin proteins are a considerable battery of large extracellular carbohydrate-active enzymes (supplemental Fig. S1), including the hyaluronidase $\mu$-toxin, which destroys the polysaccharide hyaluronan, and the large sialidases (NanJ and NanL), which remove terminal sialic acid sugars and enhance the lethal cytolytic phospholipase activity of the $\alpha$-toxin (5, 6, 18). The individual modules of such carbohydrate-active enzymes perform a variety of functions, which are most commonly catalysis, protein-carbohydrate interactions, or protein-protein interactions, and these contribute to the overall function and efficiency of the protein. However, because of their large size and flexibility, properties that make them recalcitrant to structural analyses, and their multifunctional nature, such large proteins have required a reductionist approach to their study whereby individual modules are heterologously produced for thorough structural and functional analysis. This has provided considerable insight into the individual functions of the modules. However, how the functions of the modules are spatially coordinated to cooperatively influence the overall activity of the full-length enzymes, knowledge that would have considerable biological relevance, is largely unknown as the complete structures of large carbohydrate-active bacterial virulence factors, for the most part, resist determination.

The considerable number of large and multimodular carbohydrate-active enzymes produced by C. perfringens makes this organism an excellent and relevant model system for the study of complex carbohydrate-active enzymes involved in bacterial pathogenesis. GH84C (NagJ) from C. perfringens is a homolog...
rh of the μ-toxin that comprises 1001 amino acids and four distinct modules: an N-terminal catalytic module followed by a family 32 carbohydrate-binding module (CBM32), a cohesin module (Coh), and a C-terminal fibronectin type III module (FN3) (Fig. 1). The structural and biochemical properties of the isolated catalytic module, CBM32, and Coh have been described separately (13–15, 19). These informative studies have revealed that the catalytic module is similar to human O-GlcNAcase and functions as an exo-β-D-N-acetylglucosaminidase, whereas the CBM32 preferentially recognizes the non-reducing terminus of N-acetyllactosamine (β-D-galactosyl-1,4-β-D-N-acetylglucosamine) bearing carbohydrate receptors (15, 19). The Coh module is perhaps the most unique component, because it functions to recognize and bind ultratightly to dockerin modules (Doc), such as that present at the C terminus of the μ-toxin, and plays a role in forming higher order complexes with other large C. perfringens exo-toxins, all of which are thought to contribute to the virulence of this bacterium (13). However, the overall three-dimensional structure and therefore spatial distribution of the cognate modules of GH84C remain unknown, as do the structure and function of the FN3 module. This information would provide considerable insight into the structural coordination of modular functions in this protein and provide a powerful model for understanding modularity in the carbohydrate-active enzymes of C. perfringens as well as carbohydrate-active virulence factors in other bacterial pathogens, which also frequently display modular complexity.

Intact GH84C proved intractable to heterologous production and in vitro analysis. However, a less conventional approach that combined x-ray crystallography and small-angle x-ray scattering (SAXS) was applied in a “dissect-and-build” approach to generate a solution structure model of the complete enzyme. The reconstructed architecture of GH84C in complex with a Doc-containing fragment of the μ-toxin reveals unprecedented insight into how the recognition of substrate by the catalytic module, interaction of a CBM with a carbohydrate receptor, and recruitment of other carbohydrate-active enzymes through a high affinity protein-protein interaction are spatially coordinated in an enzyme involved in a host-pathogen interaction. Furthermore, the methodology employed here should be broadly applicable to a range of large and flexible multimodular proteins.

**EXPERIMENTAL PROCEDURES**

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

5 The abbreviations used are: CBM32, family 32 carbohydrate-binding module; Coh, cohesin; FN3, fibronectin type III; Doc, dockerin; SAXS, small-angle x-ray scattering; FIVAR, found in various architectural regions module.
Structure of a Multimodular α-N-Acetylglucosaminidase

wARP (30) to build a virtually complete model with docked side chains, which was then manually completed using COOT and refinement with REFMAC. Waters were added using the ARP/wARP option in REFMAC.

The GH84C-CBM32 modular pair structure was solved by molecular replacement by first running MOLREP (31) using the GH84C catalytic module as a search model to find the single molecule in the asymmetric unit. MOLREP was run a second time to place the CBM32 model (PDB accession 2J7M) in the asymmetric unit. Successive rounds of model correction were done using COOT and refinement with REFMAC. A limited number of waters was added manually. Due to substantial disorder in portions of this structure rather than use a bulk B-factor, we opted to refine individual B-factors.

The structure of the Coh-FN3 modular pair was determined by molecular replacement using Phaser and the coordinates of the isolated Coh module (PDB accession 2O4E) as a search model to find the single molecule of Coh-FN3 in the asymmetric unit. Although the molecular replacement solution contained only ~50% of the asymmetric unit contents, the initial phases from restrained refinement with REFMAC were of sufficient quality for ARP/wARP to build a complete model, including the FN3 module, with docked side chains. Model correction was done manually using COOT and refinement with REFMAC. Waters were added using the ARP/wARP option in REFMAC.

In all cases, 5% of the reflections were flagged as “free” to monitor refinement procedures and judge model quality (32). Model validation was performed with SFCHECK (33) and PROCHECK (34). All model statistics are shown in Table 1. The coordinates and structure factors for the GH84C catalytic module, GH84C-CBM32, and Coh-FN3 have been deposited with the pdb codes of 2v5c, 2v5d, and 2w1n, respectively.

SAXS—SAXS data were collected at the X33 beamline of the European Molecular Biology Laboratory (Deutsches Elektronen Synchrotron, Hamburg) using an MAR345 image plate detector or a Pilatus 500K detector. A 4.2 mg/ml solution of bovine serum albumin was measured as a reference and for calibration. The scattering patterns were measured with an exposure time of 2 min at 288 K. The wavelength was 1.5 Å. The sample-to-detector distance was set at 2.4 m, leading to a scatter-calibration. The scattering patterns were measured with an bovine serum albumin was measured as a reference and for detector or a Pilatus 500K detector. A 4.2 mg/ml solution of bivalent module, GH84C-CBM32, and Coh-FN3 have been deposited with the pdb codes of 2v5c, 2v5d, and 2w1n, respectively.

RESULTS

Structure of the GH84C Catalytic Module—To determine the overall architecture of the enzyme we dissected GH84C into a series of modular combinations. These constructs were successfully overproduced in Escherichia coli and purified in sufficient quantities for structural analysis by x-ray crystallography and SAXS (see “Experimental Procedures” for details).

The catalytic module construct of GH84C comprising the residues 41–624 of the full protein sequence was crystallized, and its structure was determined by x-ray crystal diffraction to 2.1 Å (Table 1 and Fig. 2A). This fragment of GH84C comprises three domains that are not distinguishable through amino acid-based sequence comparisons. The N-terminal domain (amino acids 41–177) is a mixed β-sheet structure consisting of six β-strands sandwiched between three α-helices; two helices on one face of the β-sheet and one helix on the other. A central domain (amino acids 178–470) adopts a β/α barrel lacking the 7th helix, whereas the C-terminal domain (amino acids 471–624) is an elongated five-helix bundle (Fig. 2A). This is identical to the structure of a GH84C fragment from C. perfringens strain 13 reported by Rao et al. to 2.25 Å, and thus the additional structural and functional properties of this catalytic region have been discussed in detail previously (19).

Positioning the Carbohydrate-binding Module—A GH84C fragment comprising the catalytic module and the adjacent CBM32, herein called GH84C-CBM32, was also crystallized, and its structure was determined by x-ray crystallography to 3.3-Å resolution (Table 1 and Fig. 2B). Although these data were measured to comparatively low resolution the availability of the high resolution structures of the isolated catalytic module and CBM32, the latter previously determined at resolutions as high as 1.4 Å (15), made the placement of the modules by molecular replacement relatively facile and accurate. Despite
Structure of a Multimodular β-N-Acetylglucosaminidase

TABLE 1

X-ray crystal diffraction data collection and refinement statistics

|                  | GH84C-CBM | GH84C catalytic | Coh-FN3 |
|------------------|-----------|----------------|---------|
| **Data collection** |           |                |         |
| Space group       | P2₁       | I2₁2₁2₁        | C2      |
| Cell dimensions   |           |                |         |
| a, b, c (Å)       | 46.85, 69.54, 136.78 | 130.39, 150.05, 155.43 | 148.25, 48.64, 36.90 |
| α, β, γ (°)       | 90.00, 95.78, 90.00 | 20.00-1.80 (1.86-1.80) | 90.00, 96.19, 90.00 |
| Resolution (Å)    | 20.00-3.30 (3.39-3.30) | 20.00-2.10 (2.15-2.10) | 20.00-1.80 (1.86-1.80) |
| Rmerge            | 0.114 (0.287) | 0.103 (0.397) | 0.030 (0.170) |
| Rfree             | 4.4 (2.2) | 12.4 (4.8) | 30.0 (8.0) |
| Completeness (%)  | 92.8 (94.7) | 99.9 (100.0) | 95.7 (77.9) |
| Redundancy        | 2.5 (2.5) | 16.1 (15.7) | 7.1 (6.1) |
| **Refinement**    |           |                |         |
| No. reflections   | 31,497    | 1,384,042      | 166,316 |
| Unique reflections | 12,420    | 85,965         | 23,338 |
| Rmerge/Rfree      | 0.325/0.370 | 0.198/0.255 | 0.224/0.294 |
| No. atoms         |           |                |         |
| Protein           | 5,677     | 4,644 (A); 4,632 (B) | 1,750 |
| Ligands           | 3 (Ca)   | 4 (Ca); 3 (Na); 25 (CAC) | 4 (ACT) |
| Water             | 28       | 1,195          | 318    |
| B-factors (Å²)    | 52.8      | 29.3 (A); 32.1 (B) | 28.3    |
| Protein           | 71.8 (Ca) | 44.3 (Ca); 22.0 (Na); 71.7 (CAC) | 30.2 (ACT) |
| Ligands           | 27.5     | 34.3           | 38.7   |
| Water             | 0.006    | 0.010          | 0.018  |
| Bond angles (°)   | 0.934     | 1.389          | 1.732  |
| Ramachandran statistics (%) |          |                |         |
| Favored           | 84.1      | 90.4           | 91.6   |
| Allowed           | 15.5      | 9.7            | 8.4    |
| Disallowed        | 0.5       | 0.0            | 0.0    |

* Values in parentheses are for the highest resolution shells.
** All F > 2σF.

FIGURE 2. Structures of GH84C catalytic module and GH84C-CBM32 as determined using x-ray crystallography and SAXS. A and B show the crystal structures of GH84C catalytic module and GH84C-CBM32, respectively, in a ribbon representation. The arrow in B shows the C terminus of the CBM. C shows the GASBOR-generated SAXS envelope of GH84C-CBM32, whereas D shows the modules of GH84C-CBM32 manually fit into the SAXS envelope. E shows the model in D without the SAXS form. F shows the unmodified x-ray crystal structure, shown in B, fit into the SAXS-generated envelope. All of the structures are shown from identical orientations. The N-terminal domain is pictured in light blue, the catalytic TIM barrel is in orange, the helical bundle is in pale green, and the CBM in red.

the lack of excessively high mosaicity (<1) or anisotropy in the data, the refined R-factors for this structure remained relatively high at over 30%. This structure displayed substantial disorder in the N-terminal domain and portions of the (α/β)₈ barrel. Indeed, the Wilson B-factor for this data were 62 Å², which was roughly in keeping with the average refined B-factor of 52 Å² for the protein; however, the B-factors for the N-terminal domain and portions of the (α/β)₈ barrel approached 65–75 Å² indicating their relative disorder. Thus, we primarily attributed the high R-factors to this disorder, which made it difficult to accurately model portions of the structure. This structure is important, however, not for the accurate appointment of individual atoms but for its disclosure of the relative placement of the protein’s domains. The GH84C-CBM32 structure reveals the positioning of the CBM32 at the extremity of the rigid α-helical linker domain with the long axis of the CBM32 at roughly right angles to the axis of the α-helical bundle (Fig. 2B). The α-helical bundle and the CBM32 are separated by a short linker, which was clearly visible in electron density maps thus allowing unambiguous assignment of the CBM32 to the correct catalytic module in the crystal lattice. The C terminus of the CBM32 is positioned approximately at the tip of the α-helical bundle where, in the intact enzyme, the Coh module would immediately begin (Fig. 2B).

The solution conformation of the GH84C-CBM32 modular pair was also analyzed by SAXS (Table 2). The ab initio calculation of SAXS molecular envelopes for GH84C-CBM32 consistently yielded extended forms comprising a large globular region and a smaller region (Fig. 2C). An atomic model of GH84C-CBM32 based on the SAXS data was derived by two methods. In the first method, the coordinates of the composite modules were individually placed in the ab initio SAXS envelope with an attempt at minimizing bias toward the positioning of the modules in the crystal structure (Fig. 2, D and E). The crystal structure of the GH84C catalytic module was first placed...
Structure of a Multimodular β-N-Acetylglucosaminidase

in the SAXS envelope followed by rotation and translation of its domains to better fit the envelope. The coordinates of the CBM32 was then placed in the envelope in relation to the catalytic module. The relative positions of the N and C termini of contiguous domains were used as a constraint for placement. This produced a model with acceptable geometry in terms of the relative placement of the modules and gave a satisfactory χ\textsubscript{crys}\textsubscript{ol} value of 2.0. Given the structural similarity of this model to the GH84C-CBM32 crystal structure, we also modeled the entire GH84C-CBM32 crystal structure into the \textit{ab initio} SAXS envelope as a rigid body (Fig. 2F). The χ\textsubscript{crys}\textsubscript{ol} value of the model generated in this manner was 2.1 and thus not substantially different from the approach where the modules were positioned individually. The similarity of the unbiased SAXS-derived model to the crystal structure and the goodness of the fit of the unmodified crystal structure to the SAXS data indicate that the GH84C-CBM32 crystal structure is indeed representative of the solution conformation of this polypeptide.

Orientation of the Cohesin Module—Modular constructs of GH84C larger than GH84C-CBM32 proved either recalcitrant to crystallization (GH84C-CBM32-Coh) or to heterologous production (full-length GH84C). Thus, we utilized a strictly SAXS-based approach to determine the structure of GH84C-CBM32-Coh and its complex with a fragment of the μ-toxin comprising its C-terminal Doc module and the single FIVAR module (found in various architectural regions) that precedes it, hereafter called FIVAR-Doc (see supplemental Fig. S1). These studies were initiated with the CBM32-Coh construct to provide a means of determining the relative orientations of the CBM32 and Coh modules. The maximum SAXS-derived dimension of this construct, as determined by the \(D_{\text{max}}\) was ~85 Å (Table 2), which was somewhat smaller than the sum of the known maximum lengths of the isolated CBM32 (49 Å) and Coh (56 Å) modules determined from their crystal structures. The structure of this modular pair, determined by modeling the two modules into the \textit{ab initio} SAXS envelope using the positions of the N and C termi as a constraint, supports an end-to-end arrangement with an angle of ~115° between the two modules (Fig. 3A). The angling of the two modules creates a more compact structure, which explains the \(D_{\text{max}}\) that is less than the sum of the end-to-end sizes of the modules. The relatively high χ\textsubscript{crys}\textsubscript{ol} value of 2.9 for this model suggests some flexibility between these two modules that has not been accounted for in this single representative conformation. Consistent with this observation was the CBM32-Coh polypeptide’s complete recalcitrance to crystallization, implying structural heterogeneity and/or flexibility that prevented packing in a crystalline lattice. Similar flexibility of bimodular carbohydrate-active enzymes has previously been observed for cellulosomal components (43). The \(D_{\text{max}}\) for the CBM32-Coh-FIVAR-Doc complex was ~95 Å and thus only slightly more extended than the CBM32-Coh structure (Table 2). An atomic model for this complex was generated by fitting the known high resolution x-ray crystal structure of the Coh-FIVAR-Doc complex (13) as a rigid body into the \textit{ab initio} SAXS envelope followed by placement of the CBM32 module, again using the positions of the N and C termi as constraints (Fig. 3B). The modeled complex fit well into the envelope as judged by a χ\textsubscript{crys}\textsubscript{ol} value of 1.2 and perhaps reflected a more structured relative arrangement of CBM32 and Coh modules induced by the binding of the FIVAR-Doc component.

An initial putative model of GH84C-CBM32-Coh was generated by overlaying the CBM32 of the CBM32-Coh modular pair and CBM32-Coh-FIVAR-Doc SAXS models with the CBM32 of the GH84C-CBM32 crystal structure; the orientations of Coh and Coh-FIVAR-Doc portions of the model were maintained relative to the CBM32 but placed to minimize clashes with the helical bundle of the GH84C catalytic module (Fig. 3, C and D). The two resulting models, with the FIVAR-Doc portion removed from the second model, agreed well with SAXS data collected on GH84C-CBM32-Coh with χ\textsubscript{crys}\textsubscript{ol} values of 1.5 and 1.6, respectively. Indeed, the measured \(D_{\text{max}}\) of ~116 Å was consistent with the maximum dimensions of 124 Å for both of the two models (not considering the modeled FIVAR-Doc portions) (Table 2). To provide additional support for this modular arrangement the \textit{ab initio} SAXS envelope for this polypeptide was also calculated using GASBOR. The GH84C catalytic module was modeled as a rigid body, whereas the CBM32 and Coh modules were rotated and translated relative to this component to best fit the envelope. The GH84C-CBM32 crystal structure was used as a guide for the placement of the CBM32, and, as above, the positions of the N and C termi of the component modules were used to constrain the potential orientations and spatial separation of the modules. The resulting model agreed very well with the scattering data and yielded a χ\textsubscript{crys}\textsubscript{ol} of 1.2.

### Table 2

| Construct (module boundaries\textsuperscript{a}) | Molecular mass\textsuperscript{b} (kDa) | Number of amino acids\textsuperscript{c} | \(R_g\) (Å) | \(D_{\text{max}}\) (Å) | \(d_a\) (GASB) | Normalized spatial discrepancy |
|---------------------------------------------|-------------------------------|-------------------------------|----------------|-----------------|----------------|-------------------------|
| GH84C-CBM32-Coh (31–909)                  | 100                           | 902                           | 37 ± 1         | 116 ± 3         | 1.0            | 1.3                     |
| GH84C-CBM32 (31–767)                     | 82.3                          | 760                           | 35 ± 6         | 103 ± 6         | 2.3            | 1.3                     |
| CBM32-Coh (625–909)                      | 32.4                          | 307                           | 26 ± 3         | 85 ± 3          | 1.7            | 0.9                     |
| GH84C-CBM32-Coh (185–909)                | 47.9                          | 447                           | 31 ± 3         | 95 ± 3          | 1.4            | 1.2                     |
| CBM32-Coh (625–909)                      | 27.0                          | 246                           | 37 ± 6         | 118 ± 6         | ND\textsuperscript{d} | ND                     |
| Bovine serum albumin                     | 66.3                          | 582                           | 31 ± 1         | 90 ± 2          | ND\textsuperscript{d} | ND                     |

\textsuperscript{a} Construct boundaries are shown relative to the amino acid numbering in the intact GH84C protein.

\textsuperscript{b} Values are calculated including the N- or C-terminal six-histidine tags.

\textsuperscript{c} NA, not applicable.

\textsuperscript{d} ND, not determined.
Generating a Complete Model of GH84C—The structural studies of the various GH84C modules were completed by analysis of the Coh-FN3 modular pair, whose crystal structure was determined at a resolution of 1.8 Å (Table 1 and Fig. 4A). The Coh module adopts a β-sandwich fold comprising a β-sheet of four anti-parallel β-strands packing against a β-sheet of five anti-parallel β-strands, identical to that previously described for the module in isolation (13, 14). The FN3 module also adopts a β-sandwich fold of three anti-parallel β-strands atop four anti-parallel β-strands, which is characteristic of this family of modules (44–46). Overall, the crystal structure reveals an elongated structure (109 Å long) with an extended eight-residue linker separating the two modules (Fig. 4A). SAXS analysis of this modular pair yielded a $D_{\text{max}}$ of ~115 Å, consistent with the crystal structure and indicative of this polypeptide’s tendency to adopt an extended structure (Table 2). However, standard ab initio determination of SAXS envelopes yielded inconsistent results, suggesting there may be considerable conformational flexibility between the modules, as one might expect from the inter-modular linker.

Incorporation of the Coh-FN3 crystal structure into the crystal/SAXS-based structure of GH84C-CBM32-Coh enabled the generation of a complete composite model of GH84C (Fig. 5). The Coh module of the Coh-FN3 structure was overlapped with the Coh module of the GH84C-CBM32-Coh model with the lowest $\chi^2_{\text{crystol}}$ value to provide an indication of the placement of the FN3 module. The Coh module of this model was then overlapped with the high resolution crystal structure of the Coh-FIVAR-Doc complex to incorporate the non-covalently associated FIVAR-Doc modules in the complete model. The resulting overall model clearly shows the splayed architecture of this complex (Fig. 5). However, this model represents only one potential conformation of the full-length enzyme. The low resolution of the SAXS analyses made it impossible to determine the absolute relative orientations of the CBM32 and Coh modules. Although the relative angles between the individual long axes of the modules could be approximated, their relative rotational orientations around these axes could not. Therefore, assuming the CBM32 is appropriately positioned based on the x-ray crystal structure of GH84C-CBM32, this leaves some uncertainty as to the exact relative positioning of the Doc-binding site on the Coh module. However, given the steric constraints from the closely neighboring CBM32, other potential orientations of the Coh module would likely have the Doc-binding site rotated away from the CBM32 (Fig. 5).

Indeed, considering the complex structural architecture of this enzyme and the somewhat limited potential for intermodular contacts, there is likely some capacity for movement between the modules, such as that suggested by the SAXS analysis of the CBM32-Coh structure. Inter-modular flexibility is likely to be particularly pronounced within the
Structure of a Multimodular β-N-Acetylglucosaminidase

linker between the Coh and FN3. Thus, the FN3 in this enzyme likely samples a large number of spatial conformations relative to the remainder of the protein (Fig. 5).

DISCUSSION

The overall composite structure of GH84C is exceptionally revealing of how the known functions of the individual modules of GH84C are coordinated in three-dimensional space. The most immediate and striking observation is the relative positioning of the catalytic site and the binding site of the CBM32. GH84C is an exo-β-d-N-acetylglucosaminidase whose activity is likely directed toward terminal N-acetylglucosamine residues presented by host tissues (19). CBM32 binds carbohydrates bearing non-reducing terminal N-acetylgalactosamine disaccharide motifs but also binds other terminally galactosylated sugars (15). It is widely maintained that the CBMs present in carbohydrate-active enzymes target and hold it in proximity to substrate (47, 48). However, there are surprisingly few intact structures of carbohydrate-active enzymes, and no solution structures, that provide insight into how this adherence and catalysis might be coordinated. Our model of GH84C, which is based on both the crystal and solution conformations of the protein, reveals that the active site of the catalytic module and the carbohydrate-binding site of the CBM32, separated by ~60 Å, are located on the same side of the protein and are oriented...
in the same direction (Fig. 5). This suggests an optimal arrangement to simultaneously interact with and hydrolyze terminal glycans presented on a surface, such as those clustered on the outer surface of a host cell or mucosal surface. Although such a coordinated attack of binding and catalysis has been implicated in CBM-containing carbohydrate-active enzymes, this suggestion has been mainly limited to enzymes that hydrolyze insoluble plant cell-wall polysaccharides and has not been directly supported by solution structure data. The available structures of multimodular carbohydrate-active enzymes involved in bacterial pathogenesis have been primarily restricted to the sialidases from Vibrio cholerae (VCNA) and, more recently, Streptococcus pneumoniae (NanB) (49, 50). These proteins both contain a β-propeller catalytic module, which is fused to two CBM40 modules in VCNA and one CBM40 in NanB. In both of these proteins, the sialic acid-binding site of the CBM is oriented in the same direction as the catalytic site, as in GH84C; however, the distance between catalytic sites and carbohydrate binding sites is only ~25–30 Å. Although these multimodular enzymes lack the complexity typical of extracellular carbohydrate-active enzymes from bacterial pathogens, the arrangement of their catalytic sites and CBM carbohydrate-binding sites do support a model of coordinated catalysis and adherence.

One of the most fascinating and newly discovered features of C. perfringens enzymes is their capacity to interact with each other via an ultra-tight Coh-Doc interaction (13). The model of GH84C shows that the Doc-binding site on the Coh module is spatially removed from the carbohydrate-recognizing components of the enzyme thus allowing simultaneous substrate recognition and recruitment of other carbohydrate-active enzymes through the Coh-Doc interaction. Nevertheless, the structural placement of the Coh would permit another associated enzyme to lie in the same plane as the catalytic site, as in GH84C; however, the distance between catalytic sites and carbohydrate binding sites is only ~25–30 Å. Although these multimodular enzymes lack the complexity typical of extracellular carbohydrate-active enzymes from bacterial pathogens, the arrangement of their catalytic sites and CBM carbohydrate-binding sites do support a model of coordinated catalysis and adherence.

FN3 modules are found in numerous bacterial carbohydrate-active enzymes, yet little is known about their function(s). Indeed, the function of the FN3 module of GH84C remains enigmatic. In eukaryotic systems, FN3s are involved in a variety of molecular recognition processes such as cell adhesion, cell surface hormone receptors, and cytokine receptors by facilitating protein-protein interactions (51–53). The structural opposition of the FN3 from the active site of GH84C suggests that the FN3 module does not play a role in substrate recognition (Fig. 5). However, given the historical role of FN3s in protein-protein interactions and its spatial placement in the enzyme, it is conceivable that the GH84C FN3 module, like the Coh module, may mediate recruitment of other C. perfringens enzymes through an as yet unidentified interaction. Another plausible function is suggested by the unique surface charge of the FN3 module. The tip of the FN3 module that is most distal from the body of the enzyme has a distinct patch of basic residues, which is completely conserved among C. perfringens FN3 modules (not shown), leading to a cluster of positive charge at physiological pH (Fig. 4B). The charge of this patch is complementary to the negative charge of cell walls in Gram-positive bacteria, such as C. perfringens, which is imparted by the phosphodiesters bonds of teichoic acid moieties on the exterior of the bacterium (54, 55). Considering the spatial disposition of the FN3 and its associated basic patch, FN3s may function to bind the enzyme with the bacterial cell wall via an electrostatic interaction. In this scenario, the Doc-containing enzymes, which invariably lack FN3 modules or other obvious cell-wall attachment motifs, could also be tethered to the cell wall through their interaction with Coh-containing proteins, which always contain FN3 modules (supplemental Fig. S1).

The human body provides accommodation to a greater number of bacterial cells than it has of its own cells (56, 57). The majority of niches within the body that are occupied by these bacteria, such as the mucosal surfaces of gastrointestinal tract, the genitourinary tract, and the upper respiratory tract, display enormous carbohydrate content and diversity. Not surprisingly, commensal bacteria, which are often also opportunistic pathogens, as well as other dedicated pathogens, have developed extensive carbohydrate-degrading machineries to enable their persistence in the host and to promote infection. The enzymes comprising these carbohydrate-degrading systems are often large and complex and thus poorly understood. Here, we have provided a detailed look at the complete structure of such a complex enzyme from C. perfringens. As a model system, this structure provides a first glimpse at how carbohydrate adherence via a CBM, glycosidic bond hydrolysis by a catalytic module, complex formation through a protein-protein interaction, and potentially bacterial-cell wall attachment are coordinated spatially in one enzyme. Although the numerous bacterial carbohydrate-active virulence factors do not display an identical arrangement of modules, they do invariably contain catalytic modules, which are often associated with CBMs, and frequently numerous other modules of unknown functions. The investigation of GH84C presented here provides a paradigm to guide the analysis and interpretation of other important multimodular microbial enzymes.

Acknowledgments—We thank Prof. Gideon Davies for providing the coordinates of BtGH84 prior to their deposition and the staff members of beamline X6A at the National Synchrotron Light Source for beam time and assistance. We acknowledge A. Kiklin, A. Zosulya, and D. Svergun for the use of and assistance with the X33 SAXS beamline of the European Molecular Biology Laboratory outstation, Deutsches Elektronen Synchrotron, Hamburg.

REFERENCES
1. Bryant, A. E., Bayer, C. R., Chen, R. Y., Guth, P. H., Wallace, R. J., and Stevens, D. L. (2005) J. Infect. Dis. 192, 1014–1022
2. Hynes, W. (2004) Front. Biosci. 9, 3399–3433
3. Smedley, J. G., 3rd, Fisher, D. J., Sayeed, S., Chakrabarti, G., and McClane, J. Infect. Dis. 192, 1014–1022
