Motility and Flagellar Glycosylation in *Clostridium difficile*[▽][†]

Susan M. Twine,[1] Christopher W. Reid,[1] Annie Aubry,[1] David R. McMullin,[1] Kelly M. Fulton,[1] John Austin,[2] and Susan M. Logan[1]*

Institute for Biological Sciences, National Research Council, Ottawa, ON K1A 0R6,[1] and Bureau of Microbial Hazards, HFPB, Health Canada, Sir Frederick G. Banting Research Centre, Ottawa, ON K1A 0K9,[2] Canada

In this study, intact flagellin proteins were purified from strains of *Clostridium difficile* and analyzed using quadrupole time of flight and linear ion trap mass spectrometers. Top-down studies showed the flagellin proteins to have a mass greater than that predicted from the corresponding gene sequence. These top-down studies revealed marker ions characteristic of glycan modifications. Additionally, diversity in the observed masses of glycan modifications was seen between strains. Electron transfer dissociation mass spectrometry was used to demonstrate that the glycan was attached to the flagellin protein backbone in O linkage via a HexNAc residue in all strains examined. Bioinformatic analysis of *C. difficile* genomes revealed diversity with respect to glycan biosynthesis gene content within the flagellar biosynthesis locus, likely reflected by the observed flagellar glycan diversity. In *C. difficile* strain 630, insertional inactivation of a glycosyltransferase gene (CD0240) present in all sequenced genomes resulted in an inability to produce flagellar filaments at the cell surface and only minor amounts of unmodified flagellin protein.

Clostridium difficile, a gram-positive, anaerobic, spore-forming bacterium, is an emerging opportunistic pathogen and the leading cause of antibiotic-associated diarrhea and pseudomembranous colitis in humans. The recent emergence of the hypervirulent NAP1/027 strain in hospitals of North America has resulted in increased mortality rates (18, 19). While previous reports of *C. difficile* epidemics were restricted to single institutions or wards, more recently, there appears to be a wider distribution of outbreaks (20), accompanied by increasing severity of disease as well as a significant increase in the numbers of case fatalities reported (21). The pathogen is most frequently associated with antibiotic treatment, which disrupts the gut flora, allowing *C. difficile* to colonize and multiply (16). Extensive studies have demonstrated that two toxins, TcdA and TcdB, are responsible for severe tissue damage and consequent manifestation of disease (34). Infection with *C. difficile* can lead to severe diarrhea, abdominal pain, and further complications, such as pseudomembranous colitis, inflammation, and ulceration of the lining of the intestinal wall (5, 16). Importantly, recurrence rates following treatment can be as high as 35% irrespective of the drug used in initial treatment (10, 35). The estimated incidence in Canadian hospitals ranges from 38 to 95 cases per 100,000 patients (1), while in the United States, the estimated number of cases of *C. difficile* disease exceeds 250,000/year (36), with related health care costs of $1 billion annually (16). While prevention through antibiotic stewardship and optimal management of disease is the most obvious strategy currently used, there is a great need for alternate methods of treatment.

Prior to the production and release of toxin, the organism must germinate from a recalcitrant spore form and proceed to colonize the gastrointestinal tract. This colonization process is an important first step in the disease process, whereby the organism penetrates the mucus layer and adheres to the underlying colonic epithelial cells, thereby facilitating the delivery of toxins to host cell receptors. Adhesion, an early critical step in colonization, involves a number of virulence factors, but the precise mechanisms by which bacteria adhere to the mucosa and initiate infection remain to be elucidated. Such adhesins include the flagellum (29) and the high-molecular-weight surface layer protein (6). *C. difficile* is known to express peritrichous flagella, and it has been observed that the level of adherence of flagellated strains to the mouse cecum is 10-fold higher than the level of adherence of nonflagellated strains (29).

The flagellum plays a role in the ability of bacteria to adapt to their unique biological niches. Flagella from a wide range of bacteria have been shown to be important as both colonization and virulence factors, as well as critical to biofilm formation in many species (3, 37). In recent years, a rapidly increasing body of work has described the process of flagellar glycosylation in a diverse number of bacterial species (reviewed in reference 17). The diversity of glycan structures found on these organisms from unique environments points to a novel biological role for the respective glycans, which has yet to be revealed. In some cases, it has been demonstrated that the process of flagellar glycosylation has a role in both flagellar assembly and host-pathogen interactions (17). In *Campylobacter* spp., for example, in addition to being required for flagellar assembly, flagellar glycosylation plays a role in autoagglutination properties of cells and subsequent virulence and contributes to antigenic specificity (11). The sites of glycosylation of flagellin monomers from a diverse number of bacterial species have all been shown to reside within the two surface-exposed domains (denoted D2 and D3) of the flagellin monomer when assembled within the flagellar filament (22). Structural analysis of *Salmonella enterica* flagellin has revealed that these regions are surface exposed in the assembled filament and, hence, are well

---

TABLE 1. Diversity of glycan modifications among *C. difficile* strains

Strain	Origin/type(s)	Predicted mass (Da)[a]	Accurate mass (Da)[b]	Glycan mass(es) (predicted monosaccharide compositions) (Da)[c]	% Peptide sequence coverage[d]	Sites of modification[e]
630	Epidemic type X, 012 ribotype, A$^+$ B$^+$	30,773	33,559, 33,160	398 (**203**, 98, 115)	94	S^{141}, S^{174}, T^{183} S^{188}, S^{205}
BI-1	Fluoroquinone sensitive, toxinotype 3, nonepidemic, Minneapolis, MN, 1988	30,901	35,106, 35,473, 35,841, 36,208, 37,166	509 (**203**, 146, 160), 523 (**203**, 160, 160), 890 (**203**, 146, 146, 192, 203)	88	T^{141}, S^{174}, S^{205}
BI-7	Fluoroquinone resistant, toxinotype 3, epidemic, Portland, OR, 2003	30,901	ND[f]	509, 523 (**203**, 160, 160), 890 (**203**, 146, 146, 192, 203)	ND	T^{141}, S^{174}, S^{205}
QCD32g58	Toxinotype 0, Quebec, Canada	30,901	ND	ND	ND	ND
M7465	Toxinotype 0, Quebec, Canada, 2005	30,901	35,083, 35,465, 35,849	523 (**203**, 160, 160), 904 (**203**, 160, 146, 192, 203), 890 (**203**, 146, 146, 192, 203), 509	65	T^{141}, S^{174}, S^{205}
M9349	Toxinotype 0, Quebec, Canada, 2005	30,901	33,862, 35,082, 35,463, 35,840	509, 523 (**203**, 160, 160), 890 (203, 146, 146, 192, 203)	60	T^{141}, S^{174}, S^{205}
CM-56	A$^+$ B$^+$, Manitoba, Canada, 2000, 20 cases	30,919	ND	523 (**203**, 160, 160), 904 (**203**, 160, 146, 192, 203), 890 (**203**, 146, 146, 192, 203), 509	60	T^{141}, S^{174}, S^{205}
M46846	Toxinotype 9, Quebec, Canada, 2004	30,935	ND	320, 363, 523 (**203**, 160, 160), 904 (**203**, 160, 146, 192, 203), 890 (**203**, 146, 146, 192, 203)	60	T^{141}, S^{174}, S^{205}
06Cd-130	A$^+$ B$^+$, Manitoba, Canada, 2006, 6 cases	30,803	ND	523 (**203**, 160, 160), 904 (**203**, 160, 146, 192, 203), 890 (**203**, 146, 146, 192, 203)	77	T^{141}, S^{174}, S^{205}
CM-26	A$^+$ B$^+$, Manitoba, Canada, 2000, 44 cases	30,901	ND	523 (**203**, 160, 160), 904 (**203**, 160, 146, 192, 203), 890 (**203**, 146, 146, 192, 203)	73	ND
M23257	Toxinotype 12, Quebec, Canada, 2004	ND	ND	523 (**203**, 160, 160), 904 (**203**, 160, 146, 192, 203), 890 (**203**, 146, 146, 192, 203)	72	ND

[a] Calculated from the translated DNA sequence by using Protparam (http://ca.expasy.org/tools/protparam.html).
[b] Measured by infusion into a QTOF2 mass spectrometer, as described in Materials and Methods. In many cases, insufficient flagellin was isolated for accurate mass measurements.
[c] Determined from bottom-up examination of flagellin glycopeptides. The bold number indicates the conserved 203-Da mass.
[d] Indicates the peptide sequences determined from MS/MS of FliC tryptic digests.
[e] Sites of glycan attachment, determined by electron transfer dissociation MS of FliC tryptic peptides.
[f] ND, not determined.

positioned to facilitate a myriad of extracellular interactions with either host cells or environmental substrates.

Many of the studies of bacterial flagellar glycosylation have focused upon gram-negative organisms. Of the motile gram-positive bacteria, flagellin from *Listeria monocytogenes* has been shown to be glycosylated with β-O-linked GlcNAc at up to six sites/flagellin (23). The flagellins of *Clostridium botulinum* have also been reported to be glycosylated with legionaminic or hexuronic acid derivatives (32), and preliminary evidence for glycosylation of *C. tyrobutyricum* flagellin has been reported (4). However, a functional role for glycosylation has yet to be revealed for any of these organisms. It has been reported that purified *C. difficile* flagellin monomers from various strains migrate at a molecular weight greater than that predicted from the translated DNA sequence, but flagellin monomers showed no reactivity with standard glycan staining kits (31).

In this study, we show that flagellins of *C. difficile* strain 630 as well as those from recent clinical isolates of *C. difficile* are modified with diverse O-linked glycan moieties. In addition, we have identified through mutagenesis a glycosyltransferase gene from the flagellar biosynthesis locus; it is involved in the glycosylation process and, upon inactivation, leads to loss of surface-associated flagellin protein.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The *C. difficile* isolates examined in this study are presented in Table 1. All strains examined in this study, with the exception of QCD32g58, M7465, and M9349, do not appear to be clonally related. They are from distinct outbreaks and display unique typing profiles. Strains were grown on brain heart infusion (BHI) agar media supplemented with 0.5 g liter^{-1} cysteine-HCl, 5 mg liter^{-1} hemin, 1 mg liter^{-1} vitamin K$_1$, and 1 mg liter^{-1} resazurin. Bacteria were grown under anaerobic conditions at 37°C in anaerobe chambers using Oxoid Anaerobe Paks.

Motility assays. Motility assays were performed using motility agar tubes containing BHI medium (0.175% agar). These were stab inoculated and grown anaerobically at 37°C for 48 h (30).

Flagellin protein purification. Flagellin proteins were isolated using the following procedure. Strains were grown on supplemented BHI plates under anaerobic conditions for 24 h. Bacteria were harvested in 500 μl distilled water and vortexed for 3 min before being centrifuged at 14,000 rpm in a benchtop centrifuge for 5 min. The supernatants were lyophilized and resuspended in 100 μl of distilled water.

S-layer protein extraction. Partially purified S-layer was obtained from *C. difficile* cells by resuspending growth from a single BHI plate (24 h) in 500 μl of 0.2 M glycine buffer, pH 2.2. After 5 min, the bacterial cells were removed from the glycine buffer by centrifugation and supernatant retained. The glycine buffer of this partially purified S-layer preparation was exchanged with distilled water using an Amicon Ultra 4 centrifugal filter (Millipore Corp, Billerica, MA). These S-layer samples were then used in solution enzyme digests as described below.

MS of isolated flagellins. Purified flagellin was exchanged into aqueous 0.2% formic acid (vol/vol) by using a Centricon YM-30 membrane filter (Millipore). The resulting solution was infused in a hybrid quadrupole time of flight mass spectrometer (MS) (QTOF2; Waters, Beverly, MA) at a flow rate of 0.5 μl min^{-1}. Reconstructed molecular mass profiles of intact proteins were obtained through spectral deconvolution using MaxEnt (MassLynx software; Waters, Beverly, MA). Top-down experiments were performed as described by Schirm et al. (24) by using argon collision gas with collision energies ranging from 20 to 30 V. RF lens 1 voltage was increased from 30 to 125 V in order to obtain second-generation-fragment ion spectra. Tandem MS (MS/MS) of glycan-related ions was then carried out on glycan-related fragment ions generated in the orifice/

skimmer region of the MS (nano-electrospray ionization–front-end collision-induced dissociation MS/MS).

Solution enzymatic digests. To identify the type and location of glycosylation sites, flagellin (50 to 200 μg) or S-layer was digested with trypsin (Promega, Madison, WI) at a ratio of 30:1 (protein/enzyme, vol/vol) in 50 mM ammonium bicarbonate at 37°C overnight, as described previously (32). Protein digests were analyzed by nano-liquid chromatography MS/MS (nLC-MS/MS) using either a Q-TOF Ultima hybrid quadrupole time of flight MS (Waters, Milford, MA) (32) or an LTQ XL linear ion trap MS (Thermo Fisher Scientific, Ottawa, ON, Canada) coupled to a nanoAcuity ultrahigh-pressure liquid chromatography system (Waters, Milford, MA). MS/MS spectra were acquired automatically on doubly, triply, and quadruply charged ions.

HPLC purification of tryptic peptides. Tryptic digests of flagellin were fractionated using an Agilent 1100 series high-pressure liquid chromatograph (HPLC) with a diode array detector (Agilent Technologies, Palo Alto, CA). One hundred microliters of each tryptic digest was separated using a 4.6- by 250-mm Jupiter C_{18} reverse-phase column with a Phenomenex precolumn (Security-Guard, Torrance, CA). Peptides were separated using a linear gradient of 5 to 60% acetonitrile and 0.5% formic acid over 40 min at a flow rate of 1 ml min^{-1}. A 7-μl aliquot of each fraction was retained, and the remainder was immediately evaporated to dryness and stored at −20°C. Aliquots of each fraction were screened by nLC-MS/MS using the QTOF2 to confirm the peptide contents of each HPLC fraction.

Determination of glycan linkage sites by ETD. Electron transfer dissociation (ETD) preserves delicate modifications during the fragmentation process and is ideal for identifying the linkage sites of O-glycans (8, 27). Glycopeptide-containing HPLC fractions were infused at 1 μl min^{-1} into the electrospray ionization source of an LTQ XL linear ion trap MS (Thermo Fisher Scientific, Ottawa, ON, Canada) capable of performing ETD. Initially, collision-activated dissociation MS/MS analysis was performed on the glycopeptide ions to confirm their identity. ETD was then performed using fluoranthene as the anionic reagent and with supplementary activation enabled. The ETD reaction time was adjusted for optimal fragmentation of each glycopeptide (35 ms).

Accurate mass determination of the glycan oxonium ion and glycan-related fragment ions. Glycopeptide-containing HPLC fractions were infused at 1 μl min^{-1} into the electrospray ionization source of an LTQ XL Orbitrap MS (Thermo Fisher Scientific, Waltham, MA) and the MS/MS spectrum recorded over a period of 20 min. Accurate mass determination of the glycan oxonium and glycan-related fragment ions was achieved using a number of neighboring peptide fragment ions as internal mass standards. Resolution was typically 50,000 (50% valley definition). Collision-induced dissociation MS/MS analysis was performed on the glycopeptide ions to confirm their identity.

Sequencing of *fliC* genes from clinical isolates. *fliC* from each clinical isolate was amplified by PCR from chromosomal DNA by using specific primers *fliC* 1F (ATGAGAGTTAATACAAATGTAAGTGCTTTGATAGC) and *fliC* 1R (CTATCCTAATAATTGTAAAACTCCTTGTGGTTG). PCR products were cloned into pCR2.1 (Invitrogen) and sequenced using both forward and reverse sequencing primers by using BigDye Terminator v1.1 chemistry (Applied Biosystems, Foster City, CA). Sequencing reactions were run on a 3100 genetic analyzer from Applied Biosystems (Foster City, CA).

Inactivation of *fliC* and CD0240 genes in *C. difficile* 630 by using intron retargeting. Target sites were identified for each gene by using the TargeTron gene knockout system kit (Sigma Aldrich) and mutants generated according to the method of Heap et al. (12). Briefly, one-tube splicing by overlap extension PCR was used to assemble and amplify the PCR products containing the modified sequences for intron targeting to either the *fliC* gene or the CD0240 gene. The fragments were cloned directly into pMTL007, and the targeted intron was named to indicate the site of insertion within the gene (base number) and orientation (sense [s] or antisense [a]). The plasmids were pMTL007:Cdf-*fliC*-260a and pMTL007:Cdf-*240*-864s. Each of the modified plasmids was transferred from *Escherichia coli* CA434 into *C. difficile* 630Δerm (12) by conjugation, and transformants were selected on BHI plates containing 250 μg ml^{-1} cycloserine and 15 μg ml^{-1} thiamphenicol to select for *C. difficile* containing the retargeted plasmid integrants. Single thiamphenicol-resistant colonies were resuspended in phosphate-buffered saline and plated at an appropriate dilution on BHI plates containing 2.5 μg ml^{-1} erythromycin (Erm) to select for the presence of a spliced erythromycin retrotransposition-activated selectable marker which indicates intron integration. Erm-resistant colonies were examined by PCR using flanking primers for each gene (see Table S1 in the supplemental material) with the EBS universal primer (CGAAATTAGAAACTTGCGTTCAGTAAAC) to confirm integration into each gene of interest. In addition, PCRs with Erm-specific primers were used to confirm splicing of the retrotransposition-activated selectable marker.
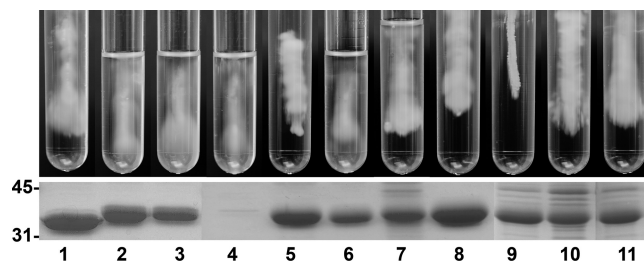

FIG. 1. Characterization of *C. difficile* isolates. The upper panel shows motility stabs of clinical isolates (0.175% agar). Motility was clearly visible in agar stabs for all strains with the exception of CM-26. The lower panel depicts an SDS-PAGE gel of FliC protein extracts. Lane 1, *C. difficile* 630; l ane 2, *C. difficile* BI-1; l ane 3, *C. difficile* BI-7; l ane 4, *C. difficile* QCD32g58; l ane 5, *C. difficile* M23257; l ane 6, *C. difficile* M46846; l ane 7, *C. difficile* M7465; l ane 8, *C. difficile* M9349; l ane 9, CM- 26; l ane 10, CM- 56; l ane 11, 06CD130. Molecular masses (in kDa) are shown to the left of the gel.

Electron microscopy. Copper grids (Electron Microscopy Sciences, Fort Washington, PA) were covered with Formvar film and coated with carbon. For sample preparation, grids were floated on a drop of bacterial cells for 5 min. After being blotted dry, the grids were negatively stained with ammonium molybdate (1%, wt/vol). Images were taken with a Zeiss EM902 transmission electron microscope operated at an accelerating voltage of 80 kV.

PCR analysis of flagellar glycosylation islands. Chromosomal DNA was prepared from clinical isolates by using the DNeasy blood and tissue kit (Qiagen Inc., Mississauga, ON, Canada). Primers spanning regions of the glycosylation island from QCD32g58 were used in reactions with chromosomal DNA and the Qiagen long-range PCR kit to determine the genetic content of each strain. PCR products were analyzed on agarose gel.

Nucleotide sequence accession numbers. The *C. difficile* fliC DNA sequences have been deposited in GenBank under accession numbers GU048823 to GU048830.

RESULTS

Motility of *C. difficile* isolates. Two recent comparative genomic studies of *C. difficile* isolates indicated that flagellum production and consequent motility may not be common features due to an absence of flagellar structural genes from a number of the genomes examined (14, 26). To determine if motility could be assessed by traditional techniques, we first examined the motility of clinical isolates by using motility agar plates. Strains of *C. difficile* were examined by stab inoculation onto standard motility agar plates (0.4% agar). In contrast to *Campylobacter jejuni*, which exhibits a diffuse spreading phenotype on this type of agar, *C. difficile* strains failed to produce the spreading pattern typical of other motile organisms (data not shown). We examined the abilities of cells from both broth- and agar-grown cultures in this manner, yet no obvious motility was observed. We next examined *C. difficile* strains stab inoculated into motility agar tubes (0.175% agar) as described by Tasteyre et al. (30). All *C. difficile* strains did display a spreading diffuse growth away from the inoculum stab in this assay, suggestive of a motile phenotype (Fig. 1). Only strain CM-26 appeared to be nonmotile by this assay.

SDS-PAGE analysis of flagellin proteins. Flagellin purified from the cell surface of *Clostridium difficile* strains (listed in Table 1) was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1). In each case, a single predominant protein band was observed in gels, and it migrated at approximately 33 to 36 kDa. Peptide MS/MS anal-
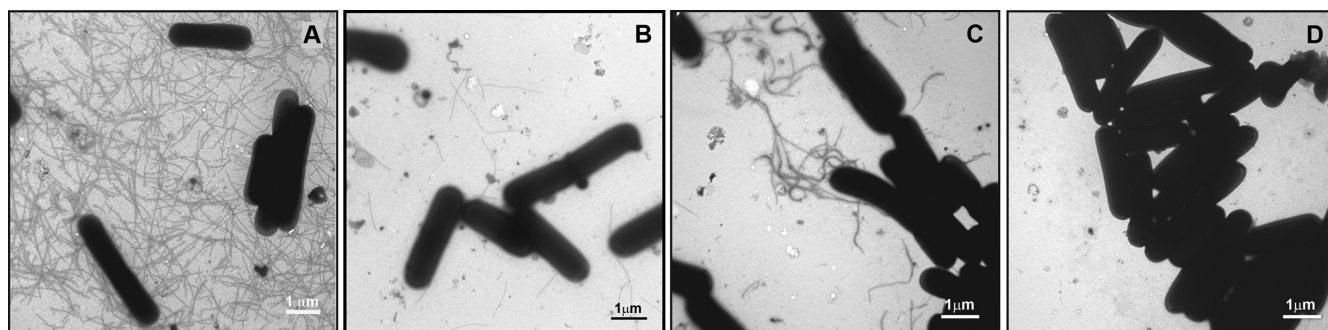
FIG. 2. Electron microscopy of uranyl acetate-stained cells. (A) *C. difficile* 630. (B) *C. difficile* QCD32g58. (C) *C. difficile* 630::0240 *erm*. (D) *C. difficile* 630::*fliC erm*.

ysis of the extracted gel band from each strain confirmed that this major protein band was the product of the *fliC* gene for each strain. Only the protein sample prepared from strain QCD32g58 failed to contain any major protein species corresponding to flagellin. However, a peptide MS/MS analysis of a tryptic digest of a partially purified S-layer preparation from this strain did identify three flagellin-specific peptides in the sample (T[39–50]AADDAAGLAISE, T[65–89]NVQDGISVVQTA EGALEETGNILQR, and T[92–107]TLSVQSSNETNTAEER), indicating that this strain does produce flagellin protein, although in much smaller amounts. Negative staining and examination of strain 630 and QCD32g58 cells by transmission electron microscopy revealed that filaments were present in QCD32g58, although in greatly reduced amounts relative to 630 (Fig. 2A and B). It should be noted that, with tube motility assays, this strain did exhibit a spreading, diffuse phenotype typical of motility (Fig. 1). In a related fashion, when strain CM-26 was tested in motility agar, it appeared nonmotile, although when cells were grown for flagellin production, examination by SDS-PAGE demonstrated that flagellin protein could be extracted from the surface of bacterial cells. Variability of both the motility phenotype and flagellin production among separate batch growths of some isolates is suggestive of a phase-variable phenotype. Recent work described phase variation of another cell surface-associated protein, CwpV (9).

Genetic analysis of flagellin genes from *Clostridium difficile* isolates. For *C. difficile* strains BI-1, BI-7, 06CD130, M9349, M46846, M7465, CM-56, and CM-26, the *fliC* structural gene was amplified from genomic DNA using PCR with flagellin-specific primers. The resulting amplicons were cloned and sequenced, and the predicted amino acid sequence and mass of each respective flagellin were determined from the DNA sequence data. With the exception of the flagellin sequence from strains M46846 and CM-56, the amino acid sequences predicted from *fliC* genes from recent clinical isolates of *C. difficile* were identical. All sequences were highly conserved, with only four amino acid substitutions in the flagellin sequence from strain M46846 and a single amino acid substitution in the flagellin sequence of strain CM-56. The complete genome sequence of strain 630 as well as shotgun sequences of eight other strains are now available through the NCBI and DOE databases. The alignment of all FliC protein sequences (see Fig. S1 in the supplemental material) demonstrated that the flagellin genes of sequenced recent clinical isolates and the

clinical isolates used in this study were remarkably conserved, with no more than seven amino acid substitutions. In contrast, *C. difficile* strain 630 showed the lowest degree of flagellin sequence conservation, with 29 amino acid substitutions compared with the other sequences. These substitutions appeared to reside mainly within the central variable region of the flagellin primary sequence. As predicted due to functional constraints for subunit-subunit interaction and proper folding of flagellin monomers, both the N- and C-terminal amino acid sequences were conserved among all strains examined (22).

Intact mass analysis of *C. difficile* flagellins. The observed masses of flagellin proteins from *C. difficile* strains 630, BI-1, M7465, and M9349 were obtained by infusion into a QTOF2 MS (Table 1). The protein mass spectrum of flagellin from each strain showed an envelope of multiply charged protein ions from which the reconstructed molecular mass profile was calculated. None of the intact masses measured matched precisely the mass predicted from the translated gene sequence, suggesting that these flagellins were posttranslationally modified. In the case of strain 630, the reconstructed molecular mass profile showed two major peaks at 33,559 and 33,160 Da (Fig. 3A), significantly greater than the predicted protein mass of 30,755 Da. Of significance, the observed major intact mass peaks were separated by a mass of ~398 Da (see below). Comparison of the observed protein intact mass of *C. difficile* 630 suggests that it corresponds to the flagellin protein FliC, bearing six or seven residues modified with a glycan of 398 Da.

The reconstructed molecular mass profile of strain BI-1 showed a more complex spectrum, with major intact mass peaks at 35,106, 35,473, and 35,841 Da. Minor peaks were clearly visible at 36,208 and 37,166 Da (Fig. 3B). All masses were greater than the mass predicted from the translated gene sequence (30,901 Da). The reconstructed molecular mass profiles of *C. difficile* strains M9349 and M7465 also showed a more complex intact molecular mass profile, which closely resembled that of BI-1, with intact masses listed in Table 1.

Top-down analyses of *C. difficile* flagellins. Previously, we have shown that examination of intact flagellar proteins by electrospray ionization MS allows the identification of labile-glycoprotein-associated ions, in particular, glycan-related ions (24). We used this technique to examine the posttranslational modifications of *C. difficile* flagellins in this study (Fig. 3C and D).

MS/MS analyses of multiply charged protein ion precursors of FliC from strain 630 showed an abundant ion at *m/z* 399.1
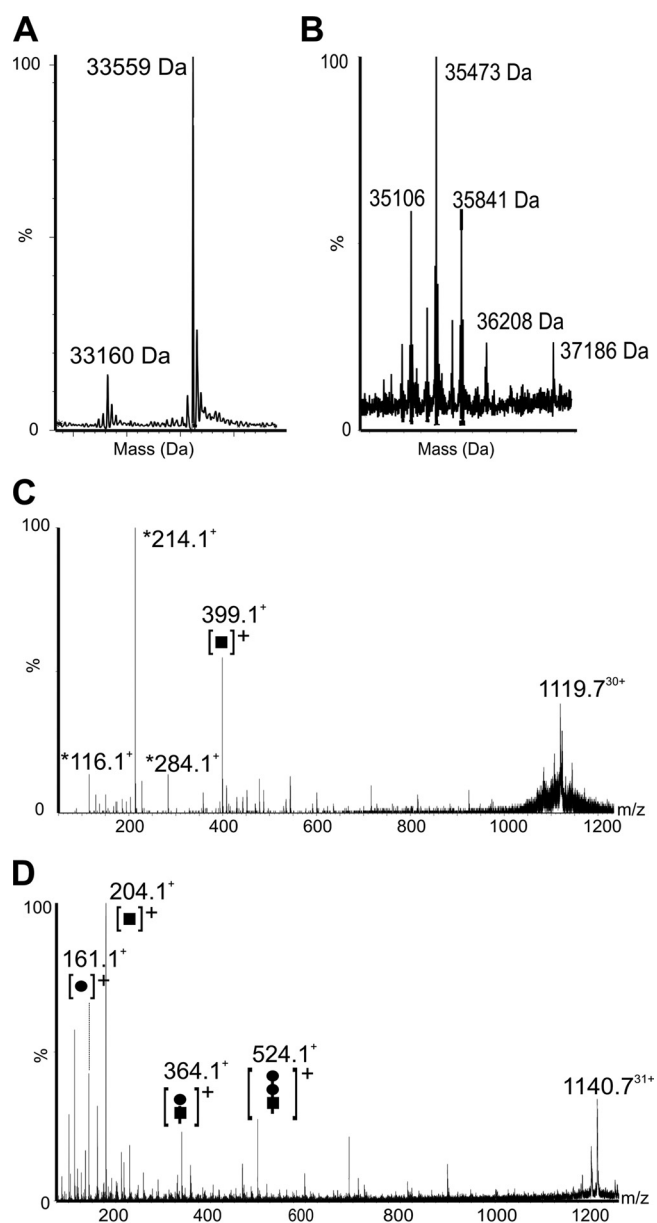
FIG. 3. Electrospray MS analyses of intact flagellin from strains of *C. difficile*. (A) The reconstructed molecular mass profile of FliC from *C. difficile* 630, showing two peaks at 33,559 and 33,160 Da. These are 2,804 and 2,405 Da larger than that predicted from the translated gene sequence of *fliC* and correspond to the attachments of six and seven glycan moieties of 398 Da. No signal corresponding to the unmodified flagellin was observed. (B) The reconstructed molecular mass profile of FliC from *C. difficile* BI-1, showing intense peaks at 35,106, 35,473, and 35,841 Da and less intense peaks at 36,208 and 37,186 Da, all larger than the mass of FliC, which was predicted to be 30,901 Da. (C) Tandem MS analysis of the multiply charged protein ion from *C. difficile* strain 630 at m/z 1,119.7^{30+}. Intense ions were observed at m/z 399.1, the putative glycan observed to modify FliC from this strain, and glycan-associated fragment ions were observed at m/z 284.1, 214.1, and 116.1 (indicated by an asterisk). (D) MS/MS analysis of the multiply charged protein ion from *C. difficile* BI-1 m/z 1,140.7^{31+}. Numerous ions were observed in the resulting MS/MS spectrum, including glycan-associated ions at m/z 524.1, 364.1, 204.1, and 161.1. MS/MS spectra were acquired at collision energy of 15 to 35 V by using argon as the collision gas.

(Fig. 3C). Increasing the voltage of RF lens 1 promoted the formation of labile-protein-associated fragments ions in the orifice/skimmer region of the MS, allowing MS/MS spectra of the ions to be recorded. Second-generation MS/MS spectra were recorded for the ion at m/z 399.1, which gave dominant fragment ions at m/z 284.1, 214.1, 186.1, and 116.1. No fragment ions corresponding to peptide type y or b ions were observed, suggesting that the observed fragment ions were more typical of those observed for glycan fragmentation.

In contrast to *C. difficile* 630 flagellin (FliC), a similar MS/MS analysis of a multiply charged protein ion of FliC from *C. difficile* strain BI-1 showed an intense ion at m/z 204.1, which would typically correspond to a *N*-acetyl hexosamine (HexNAc) residue (Fig. 3D). No evidence for a glycan oxonium ion at m/z 399 was observed for this flagellin sample. Other ions, of weaker intensity, were observed at m/z 161.1, 364.1, and 524.1. As with *C. difficile* 630 FliC, no ions characteristic of peptide b or y ions were observed, suggesting that the ion observed at m/z 524.1 was a glycan ion. Also observed in this region were fragment ions at m/z values of 186.1, 168.1, and 138.1, characteristic of HexNAc sugar fragmentation. Of significance, neutral losses of 160 Da were observed from the precursor ion, suggesting that the ion observed at m/z 161.1 is glycan related. From this initial analysis, it is apparent that the glycan present on flagellin from *C. difficile* BI-1 is quite distinct from that found on *C. difficile* 630. Similar second-generation ion spectra were recorded for multiply charged protein ions from *C. difficile* strains M9349 and M7465, with product ion spectra similar to those observed for MS/MS of multiply charged protein ions from strain BI-1 (data not shown).

nLC-MS/MS analysis of *C. difficile* flagellin peptide digests. Preliminary data were collected by analysis of the tryptic digest of the *C. difficile* 630 FliC by nLC-MS/MS. Many of the MS/MS spectra could be readily assigned to unmodified flagellin tryptic peptides. A number of spectra, however, were derived from peptides that appeared to possess a glycan modification. Figure 4A shows the peptide MS/MS spectrum of the tryptic glycopeptide T$^{176-189}$IQLVNTASIMASAGITTASIGSMK. Peptide type y and b fragment ions were observed at very low intensities, with the spectrum dominated by a glycan oxonium ion at m/z 399.1. The total mass of the glycopeptide was 3,160.24 Da (predicted peptide mass, 2,364.26 Da), which suggested modifications with two glycans of 398 Da. This glycan was also found on three other glycopeptides (T$^{135-144}$LLDGTSSTIR, T$^{202-212}$TMVSSLDAALK, and T$^{191-201}$AGGTTGTDAAK). In contrast to the case for *C. difficile* 630 FliC, bottom-up analysis of FliC from BI-1 revealed a distinct pattern of glycosylation. Figure 4B shows the MS/MS spectrum of the modified tryptic peptide from BI-1 FliC, T$^{135-144}$LLDGSSTEIR. A number of modified tryptic peptides were detected (T$^{145-166}$LQVGANFGTNVAGTTNNNNEIK, T$^{167-178}$VALVNTSSIMSK, T$^{202-212}$QMVSSLDVALK, and T$^{124-134}$ISSSIEFNGKK), all with what appeared to be a 523-Da modification, composed of residue masses of 160, 160, and 203 Da. Neutral losses of 160, 160, and 203 Da were observed in the high-m/z region of the spectrum. Weak fragment ions at m/z 204.1 and 161.1 were also observed. It was possible to infer the peptide sequence from the peptide y ion sequence. In an attempt to increase the sequence coverage, flagellin tryptic peptides from strain BI-1 were separated by HPLC. MS/MS analyses of the HPLC frac-
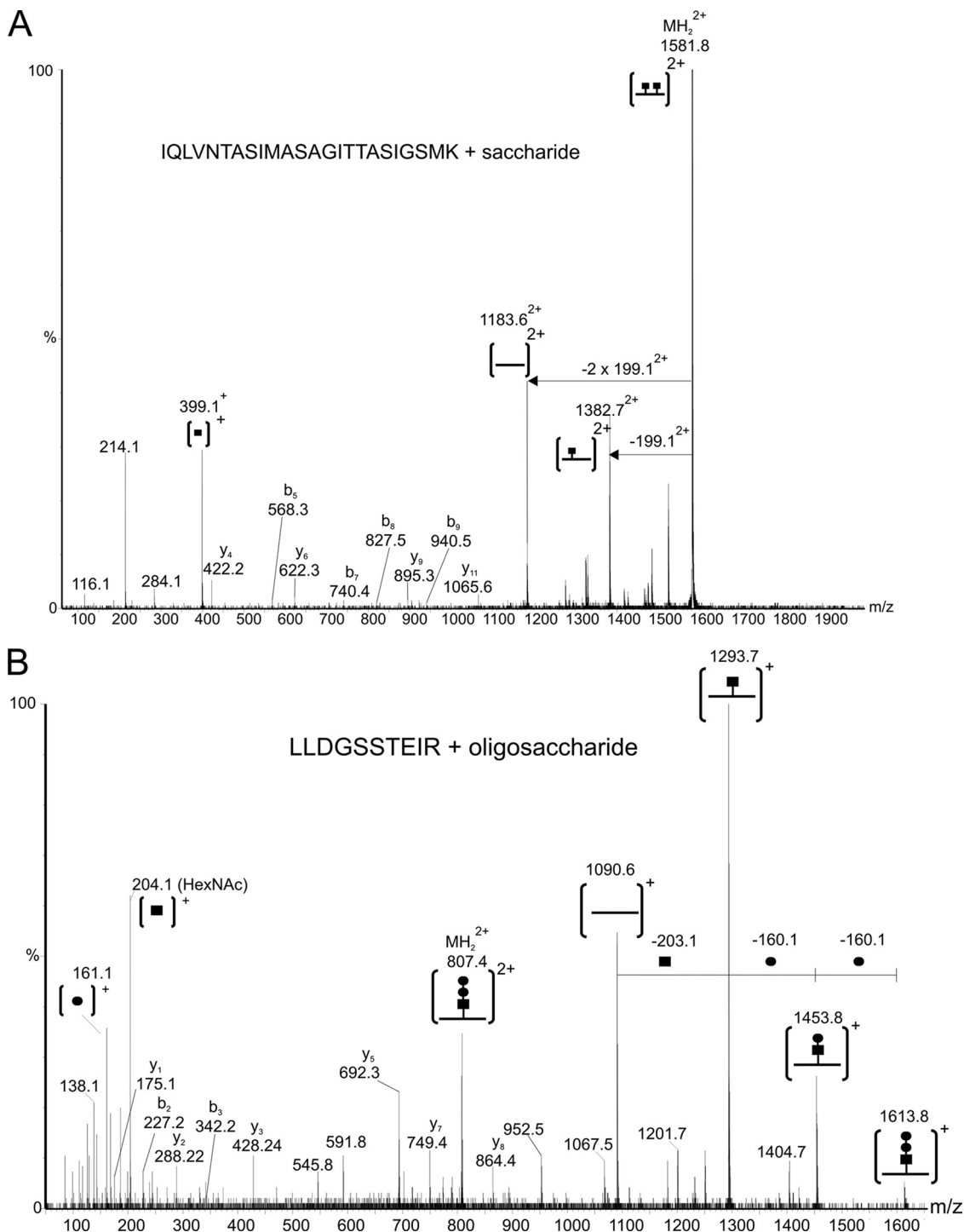
FIG. 4. MS analysis of modified tryptic peptides from flagellar proteins of *C. difficile* strains 630 and BI-1. (A) nLC-MS/MS spectrum of the doubly protonated $T^{167-190}$ glycopeptide ion at $1,581.8^{2+}$ from *C. difficile* 630. The spectrum is dominated in the high-molecular-*m/z* region by the sequential loss of two carbohydrate moieties. Two sequential losses of *m/z* 199.1^{2+} corresponds to the loss of a doubly charged glycan of a residue mass of 398.1 Da (indicated by the arrow). The low-*m/z* region of the spectrum is dominated by the carbohydrate oxonium ion (*m/z* 399.1) and what could be a fragmentation component (*m/z* 214.1). (B) nLC-MS/MS spectrum of the doubly protonated $T^{135-144}$ glycopeptide ion 807.4^{2+} from *C. difficile* BI-1. The high-*m/z* region is dominated by the sequential loss of *m/z* 160.1, 160.1, and 203.1 from what appears to be a trisaccharide. Oxonium ions are visible at *m/z* 204.1 (HexNAc) and *m/z* 161.1. The glycan appears to be O linked through the HexNAc residue.

tions showed glycopeptides with identical amino acid sequence but variable glycan modifications. For example, the tryptic glycopeptide $T^{135-144}$LLDGESTEIR was observed to be modified with a glycan of 523 Da, 904 Da, or 840 Da. Significantly,

each glycopeptide eluted at distinct retention times, indicating that these are distinct glycan modifications rather than experimental artifacts generated by in-source MS fragmentation. There appeared to be no pattern in site occupancy, with each
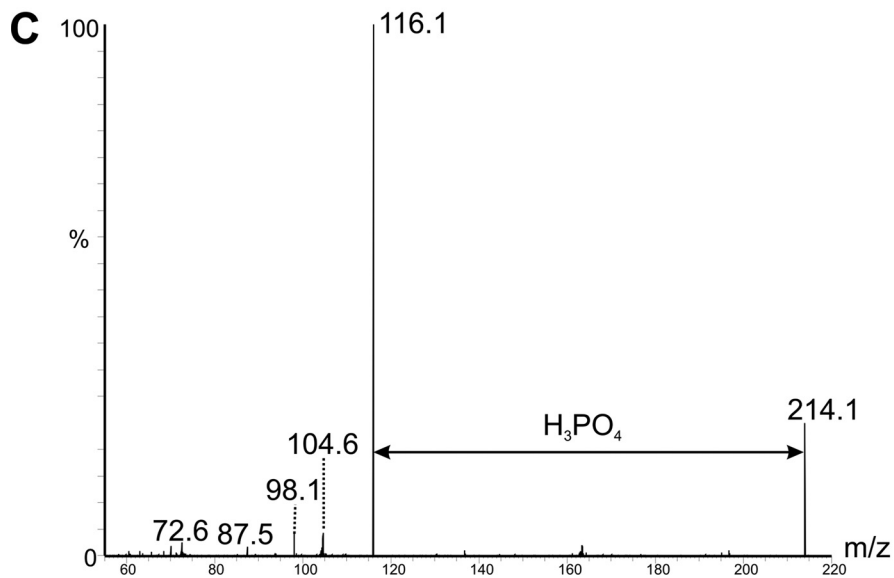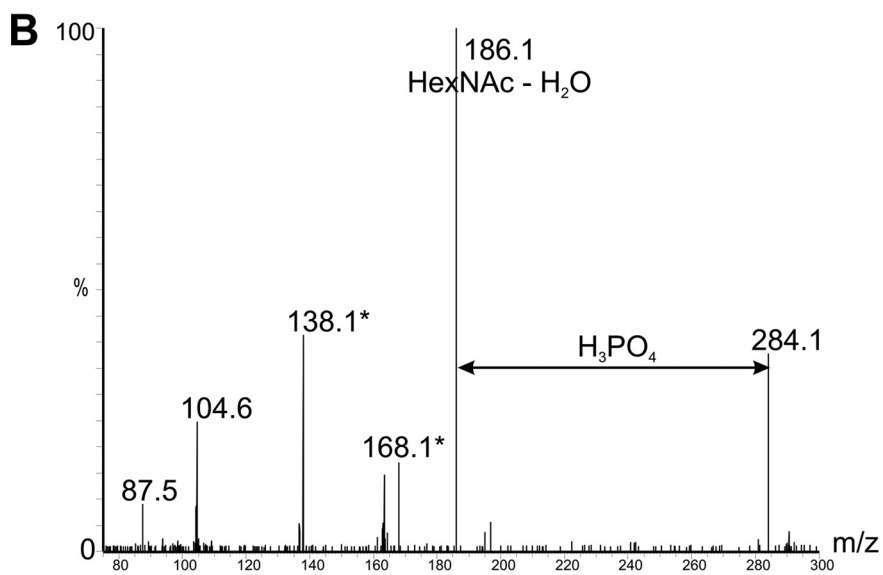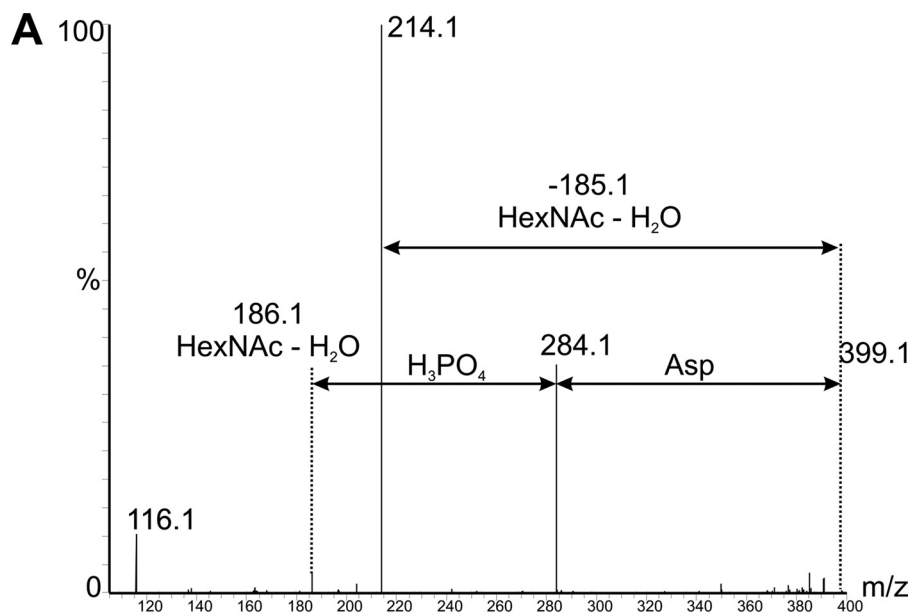
glycopeptide observed to be modified with the 523-Da, 904-Da, or 840-Da glycan. A similar pattern of glycan modification was observed for flagellins of other clinical *C. difficile* isolates. These are summarized in Table 1.

Glycopeptides were characterized from clinical strains from both Quebec and Manitoba, Canada, and shown to contain glycan structures closely resembling those found on BI-1 flagellin. Glycopeptide MS/MS spectra from BI-1 also showed a series of neutral losses of 203, 160, and 146 Da from the modified glycopeptide ion. In BI-1, BI-7, and the Canadian clinical isolates, MS fragmentation of glycopeptides revealed that HexNAc was the linking sugar connecting to the protein backbone (203.1 Da) with a oligosaccharide chain composed of putative deoxyhexoses (146 Da) or methylated deoxyhexoses (160 Da) and a monosaccharide of 192 Da (putative heptose). These were present in various combinations to form glycans of unique chain lengths (Table 1).

Determination of glycan attachment sites by ETD. Analysis of the sequence of glycopeptides did not reveal any N-linked sequons; therefore, the modification was thought to be O linked. ETD preserves delicate modifications during the fragmentation process and is ideal for identifying the linkage sites of O-linked glycans (8, 13). Glycopeptide-containing HPLC fractions of *C. difficile* 630 flagellin digest were infused into the electrospray ionization source of an LTQ XL linear ion trap capable of performing ETD. Initially, collision-activated dissociation MS/MS analysis was performed on the glycopeptide ions to confirm their identity. ETD was performed using fluoranthene as the anionic reagent and with supplementary activation enabled. The ETD reaction time was adjusted for optimal fragmentation of each glycopeptide. Five of the seven sites of glycosylation were identified in this manner for *C. difficile* 630 FliC as S^{141}, S^{174}, T^{183}, S^{188}, and S^{205}, and three of these sites were confirmed as sites of glycosylation in seven additional strains (Table 1). It appears that the O-linked sites of glycosylation are conserved among strains and are localized to the central variable region of the FliC protein, which forms the surface-exposed D2 and D3 domains within the folded monomer.

Structural characterization of unusual O-linked glycan moieties by MS/MS analyses. One of the challenges in structural elucidation of novel bacterial glycans is the relatively poor sensitivity of nuclear magnetic resonance, which requires milligram quantities of purified material. Due to the relatively low abundance of this protein on the cell surface, we have not been able to obtain a sufficient quantity of glycan for detailed structural analysis. However, high-resolution MS can provide a considerable amount of information on glycan structure and composition. Fragmentation patterns of glycans attached to protein can provide information on the organization of the glycan structure as well as identify the linking sugar moiety.

Further inspection of the *m/z* 399.1 glycan oxonium ion MS/MS spectrum from a *C. difficile* 630 flagellin glycopeptide

yielded information regarding the glycan composition. Loss of 115 Da from the parent ion yielded a fragment ion at *m/z* 284.1, and neutral loss of 98 Da from this ion produced an ion of *m/z* 186.1. The loss of 98 Da could plausibly be assigned to phosphoric acid, yielding a dehydrated HexNAc (*m/z* 186.1) (Fig. 5A). Furthermore, loss of dehydrated HexNAc (185 Da) from the *m/z* 399.1 parent ion gave rise to an intense fragment ion at *m/z* 214.1 (Fig. 5A). Further MS/MS spectra of the glycan-associated fragment ions were collected. Triple MS of *m/z* 284.1 showed a loss of phosphoric acid (98 Da), to give a fragment ion at *m/z* 186.1 (Fig. 5B). Triple MS of a glycan fragment ion at *m/z* 214.1 showed a neutral loss of 98 Da (phosphoric acid) and a strong ion at *m/z* 116.1 (Fig. 5C).

Where possible, accurate mass measurements were performed for the glycan-related fragment ions in glycopeptide MS/MS spectra, and these were used to determine plausible elemental formulae of unknown glycans. The accurate masses and top-ranked plausible elemental compositions are shown in Table 2. The top-ranked elemental formula of the ion of *m/z* 186.1 indicated this to be a dehydrated HexNAc. This was further confirmed by the presence of a weak ion at *m/z* 204.1 in several glycopeptide MS/MS spectra and characteristic HexNAc-associated fragment ions (*m/z* 168.1 and 138.1 [15]). The top-ranked elemental formula of the glycan-associated fragment ions at *m/z* 214.1 and 284.1 supported our suggestion of a phosphate linkage within the sugar. The elemental formula of the unknown fragment ion of *m/z* 116.1 suggested a methylated aspartic acid. The quadruple MS fragmentation of *m/z* 116.1 supported this, with fragment ions characteristic of an aspartic acid immonium ion (*m/z* 88.1) and amino acid backbone (*m/z* 57) (data not shown). The 398-Da flagellar glycan is likely to be a HexNAc linked to methylated aspartic acid via a phosphate linkage. The glycan is linked to the protein via the HexNAc moiety.

Flagellins from *C. difficile* strains BI-1 and BI-7 and the Canadian clinical isolates were found to be modified with several novel glycan moieties, ranging in mass from 494 Da to 917 Da (Table 1). The complex oligosaccharides are composed of various components, which were identified by their characteristic neutral losses and corresponding glycan oxonium ions. Plausible candidates, based upon residue mass for each residue, are HexNAc (203 Da), deoxyhexose (146 Da), and heptose (192 Da). Accurate mass measurements where glycan-related ions were observed in peptide MS/MS spectra, were performed. From these measurements, plausible elemental formulae were calculated for the 523-Da glycan and the *m/z* 204.1 and 161.1 glycan-related ions. The top-ranked elemental formula for *m/z* 204.1 glycan fragment ion was consistent with HexNAc, while the top-ranked elemental formula for *m/z* 161.1 fragment ion was consistent with a methylated deoxyhexose (Table 2).

Bioinformatic analysis of *C. difficile* flagellar glycosylation locus. The genome of *C. difficile* 630 was recently published

FIG. 5. Analysis of the unusual carbohydrates obtained by nano-electrospray ionization–front-end collision-induced dissociation MS/MS. (A) MS/MS spectra of the oxonium ion at *m/z* 399.1. Intense fragment ions were observed at *m/z* 284.1 and 214.1, with weaker ions at *m/z* 204.1 and 116.1. (B) MS/MS spectrum of glycan fragment ion at *m/z* 284.1. Neutral loss of 98 Da from the parent ion resulted in a fragment ion at *m/z* 186.1. (C) MS/MS spectrum of glycan fragment ion at *m/z* 214.1. A neutral loss of 98 Da resulted in an observed fragment ion at *m/z* 116.1.

TABLE 2. Accurate masses of the glycan oxonium and glycan-related fragment ions and most plausible empirical formula[a]

Carbohydrate fragment ion	Strain	Mean accurate $m/z \pm$ SD	Residue mass (Da)	Elemental composition	Predicted oxonium ion (m/z)
399.1	630	399.1164 ± 0.0012	398.1086	$C_{13}H_{24}N_2O_{10}P$	399.1163
284.1	630	284.0536 ± 0.0010	284.0458	$C_8H_{15}NPO_8$	284.0530
214.1	630	214.0481 ± 0.0072	213.0403	$C_5H_{13}NO_6P$	214.0475
204.1	630	204.0865 ± 0.0004	203.0787	$C_8H_{14}NO_5$	204.0867
186.1	630	186.0766 ± 0.0007	185.0688	$C_8H_{12}NO_4$	186.0761
116.1	630	116.0706 ± 0.0003	115.0628	$C_5H_{10}NO_2$	116.0706
524.2	BI-1	524.2311 ± 0.0049	523.2231	$C_{22}H_{38}NO_{13}$	524.2338
204.1	BI-1	204.0864 ± 0.0059	203.0745	$C_8H_{14}NO_5$	204.0866
161.1	BI-1	161.0811 ± 0.0072	160.079	$C_7H_{13}O_4$	161.0808

[a] For *C. difficile* strain 630, fragment ions y_3 (m/z 389.2513), y_2 (m/z 288.2036), and y_8 (m/z 836.4114) from the peptide T$^{135-144}$LLDGTSSTIR were used as references. Hydrogen mass, 1.0078 Da. For *C. difficile* strain BI-1, fragment ions y_3 (m/z 417.2462), y_2 (m/z 288.2036), and y_8 (m/z 807.335) from the peptide T$^{135-144}$LLDGSSTEIR glycopeptide ion 807.4^{2+} were used.

(25), and seven additional shotgun sequences of highly virulent *C. difficile* isolates from Quebec, Newfoundland, and Ontario, Canada, Minneapolis, MN, and Marne, France, are available through the DOE Joint Genome Institute integrated database (http://img.jgi.doe.gov). A preliminary BLAST analysis of the flagellin genetic locus revealed that glycan biosynthesis genes were present in all genomes in close proximity to the flagellar structural gene *fliC*. Bioinformatic analysis of the genes lying
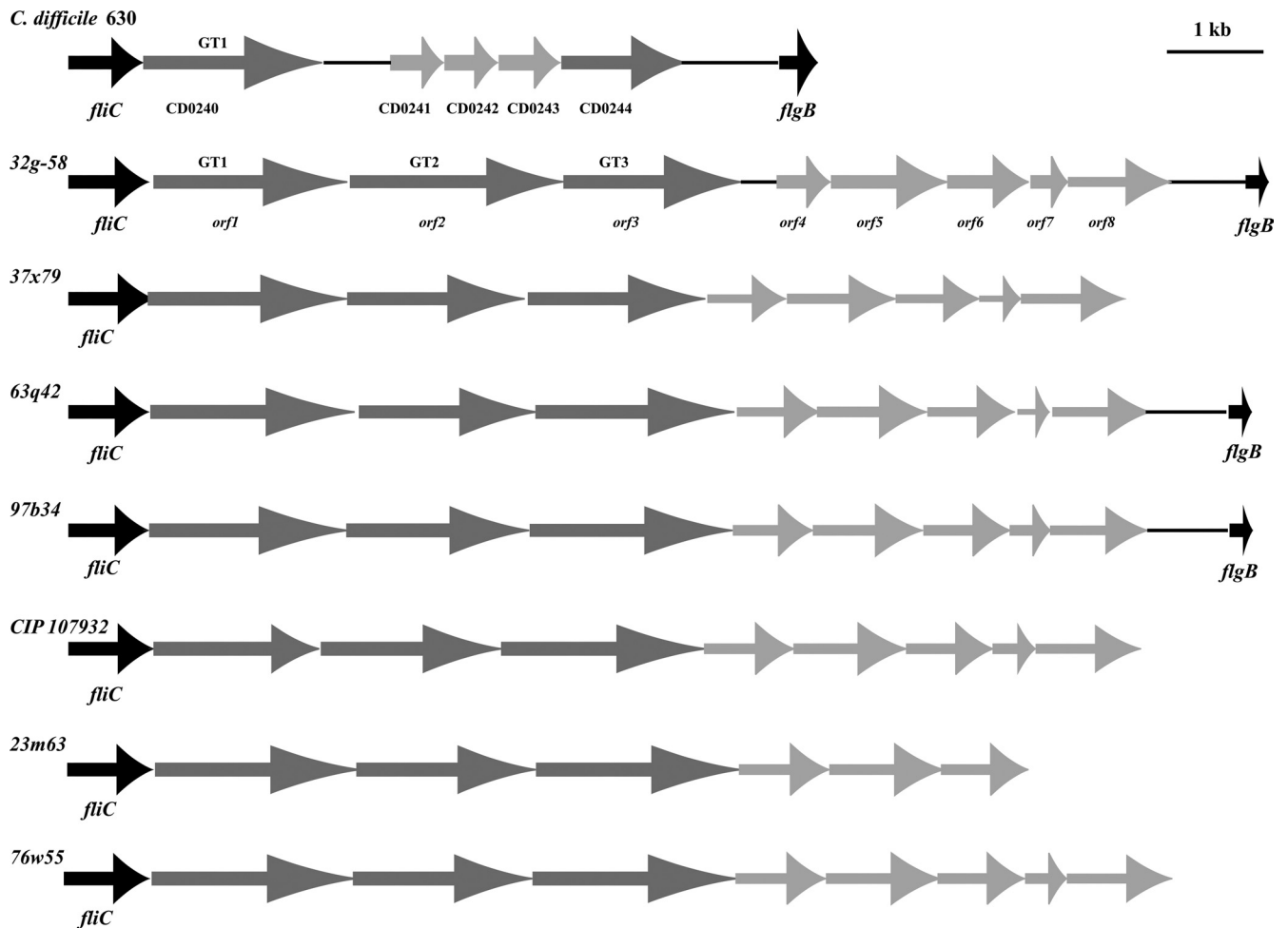

FIG. 6. Schematic diagram showing the genetic organization of the flagellar glycosylation locus in *C. difficile* strains. The diagram shows the region of the chromosome flanked by the *fliC* and *flgB* structural genes, which encode flagellin (FliC) and FlgB (rod protein), respectively. Flagellar structural genes *fliC* and *flgB* are colored black; for the flagellar glycosylation locus genes, glycosyltransferase genes are dark gray and precursor glycan biosynthesis genes are light gray. In the *C. difficile* 630 locus, the genes annotated as a phosphatase gene (CD0241), sugar nucleotide transferase gene (CD0242), and hypothetical ORF (CD0243) are colored light grey. The available shotgun sequence data for strain QCD32g58 (*orf2*/GT2 and *orf3*/GT3) and for CIP107932 (*orf1*/GT1) likely contain sequencing errors, which result in truncation of the respective ORF. The schematic displays the full-length ORF for each of these genes.

TABLE 3. Flagellar glycosylation locus genes and similarity of predicted proteins to proteins in the microbial genome database

Strain and gene	Predicted function(s) (E value)	*C. difficile* homolog(s)
630		
CD0240 (GT1)	β-(1,4)-Glycosyltransferase ($6e^{-29}$), ExoA homolog ($2e^{-12}$), GT family 2 ($5e^{-10}$)	CIP 107932, 23m63, 32g58, 37x79, 63q42, 76w55, 97b34,
CD0241	Phosphoserine-phosphatase-like protein ($2e^{17}$)	None
CD0242	Phosphocholine cytidylyltransferase ($1e^{-21}$)	None
CD0243	Methyltransferase ($2e^{-27}$)	None
CD0244	CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase ($1e^{-18}$), cholinephosphotransferase ($3e^{-7}$), GT family 2 ($3e^{-6}$)	None
QCD32g58		
orf1 (GT1)	GT family 2 ($2e^{-57}$)	37x79, 63q42, 97b34, CIP 107932, 23m63,76w55
orf2 (GT2)	GT family 2 ($3e^{-38}$), methyltransferase ($2e^{-13}$)	37x79, 63q42, 97b34, CIP 107932, 23m63,76w55
orf3 (GT3)[a]	GT family 2 ($9e^{-73}$), streptomycin biosynthesis protein ($4e^{-47}$)	QCD32g58, 63q42, 97b34, CIP 107932, 23m63,76w55
orf4	Carbamoyl-phosphate synthase large subunit ($6e^{-38}$), biotin carboxylase ($1e^{-20}$)	37x79, 63q42, 97b34, CIP 107932, 23m63,76w55
orf5	Ornithine cyclodeaminase ($2e^{-74}$)	37x79, 63q42, 97b34, CIP 107932, 23m63,76w55
orf6	FdtA, dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase ($2e^{-29}$)	37x79, 63q42, 97b34, CIP 107932, 23m63,76w55
orf7	TDP-3-aminoquinovose aldamine ($2e^{-101}$), glutamine-scyllo-inositol transaminase ($4e^{-133}$)	37x79, 63q42, 97b34, CIP 107932, 76w55
orf8	3-Amino-5-hydroxybenzoic acid synthase family ($5e^{-119}$), aminotransferase family ($1e^{-9}$)	37x79, 63q42, 97b34, CIP 107932, 76w55

[a] *orf3* protein sequence from *C. difficile* 37x39 used in Blastp, as potential sequencing errors in QCD32g58 gene cause a truncation in the ORF.

immediately downstream in *C. difficile* 630 indicated that between the *fliC* structural gene and *flgB* structural gene lies a putative flagellar glycan biosynthesis locus (CD0241 to CD0244). In the clinical isolates, this locus showed considerable genetic diversity and increased size compared to that of *C. difficile* 630 (Fig. 6). *C. difficile* 630 has the least complex complement of carbohydrate biosynthesis genes in this region. Immediately downstream of *fliC* is a putative glycosyltransferase gene (CD0240). In addition, four other genes, namely, CD0241, CD0242, CD0243, and CD0244, are present and appear to encode a putative phosphatase, a sugar nucleotide (nucleoside triphosphate) transferase, a hypothetical open reading frame (ORF), and a second glycosyltransferase, respectively, based on Blastp analysis (Table 3).

In contrast, the corresponding genetic locus in the seven shotgun-sequenced isolates is much larger and is highly conserved in genetic content (Fig. 6). Each strain has three ORFs lying immediately downstream of the *fliC* gene which are transcribed in the same orientation and which encode glycosyltransferases (GT1, GT2, and GT3). For *C. difficile* strain CIP 107932, the GT1 gene is annotated as two separate ORFs, and for *C. difficile* QCD32g58, the GT2 and GT3 genes are annotated as two separate ORFs, although the overall sequence is highly conserved among all strains (>90% identity) and it is likely that each GT is a single continuous ORF. The annotation of each GT into two ORFs is likely due to sequencing errors in these shotgun sequences. The GT2 ORF appears to encode a bifunctional protein which contains a GT2 family domain (PFAM 00535), as well as a methyltransferase domain (PFAM 08241). In addition to containing the three conserved GT ORFs, this extended locus also contains a number of ORFs encoding proteins with predicted enzymatic functions of relevance to glycan biosynthesis pathways. These include proteins with homology to a dehydrogenase, deaminase, aminotransferase, and isomerase enzyme (Table 3). A hypothetical protein with homologs in a number of bacterial species was also present in all virulent strains examined. While the genetic content of this locus is conserved, it is quite distinct from that found in the *C. difficile* 630 genome. However, a homolog of CD0240 (with >80% identity) which lies proximal to the *fliC* structural gene is present in all strains (GT1). PCR analysis using primers specific to genes of this glycosylation island with chromosomal DNA from each clinical isolate characterized in this study confirmed the presence of a homolog of the GT1/CD0240 gene in each of these strains. In addition, appropriate-sized PCR fragments were obtained when primers spanning the genes of the extended glycosylation island were used, demonstrating the conservation of this extended locus among the recent Canadian clinical isolates (data not shown).

Characterization of *fliC* and CD0240 mutant strains of *C. difficile*. To further understand the contribution of the flagellin protein to motility of *C. difficile* cells and to characterize the flagellar glycosylation pathway, we next generated insertionally inactivated *fliC* and CD0240 mutants by using the recently published ClosTron targeted mutagenesis approach (12). Insertion of the TargeTron Erm resistance marker into either *fliC* or CD0240 was confirmed by PCR using primers flanking the gene of interest and with primers specific to the TargeTron Erm resistance marker (data not shown). The mutant strains were grown overnight, and flagellin protein extracts prepared as described above were analyzed by SDS-PAGE. As can be seen in Fig. 7, production of flagellin protein in the *fliC* mutant was completely abolished in contrast to that of the parent (lane 3). In contrast, analysis of the flagellin extract from the *C.*
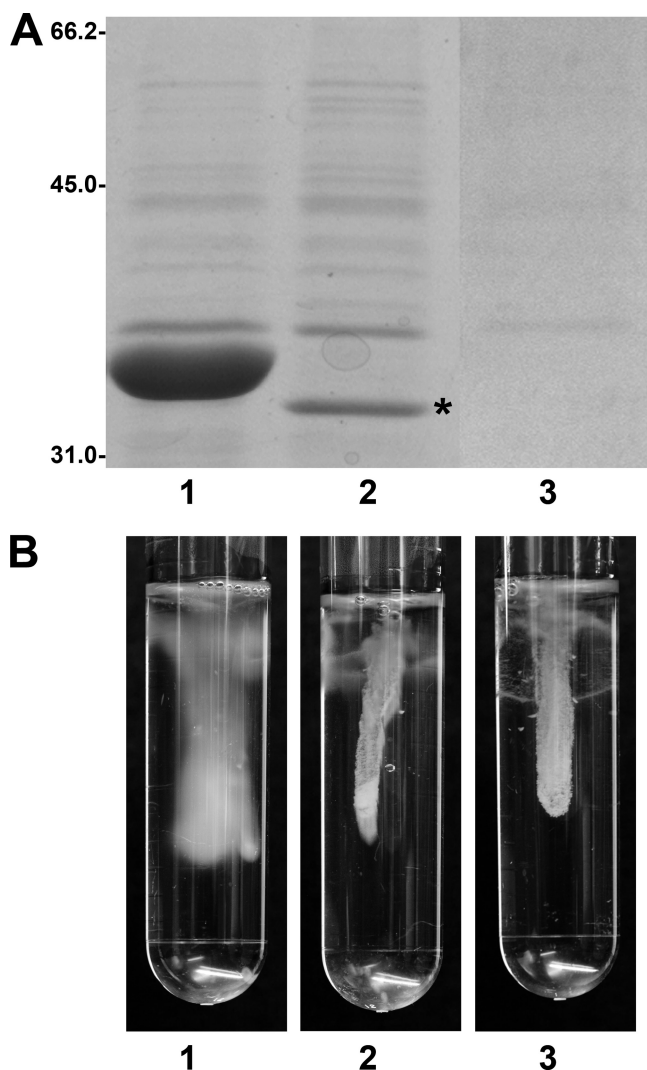
FIG. 7. Characterization of *C. difficile* mutants. (A) SDS-PAGE analysis of flagellin extracts of *C. difficile* 630 (lane 1), *C. difficile* 630::0240*erm* (lane 2), and *C. difficile* 630::*fliC*erm (lane 3). Molecular masses (in kDa) are shown to the left of the gel. Asterisk, protein of a molecular mass corresponding to unmodified flagellin. (B) Motility stabs (0.175% agar) show *C. difficile* 630 (lane 1), *C. difficile* 630::0240*erm* (lane 2), and *C. difficile* 630::*fliC*erm (lane 3).

difficile 630::0240*erm* strain revealed a minor amount of a protein of a molecular mass corresponding to unmodified flagellin (Fig. 7A, lane 2). The ability to glycosylate the flagellin protein appears to be required for optimal production at the cell surface. We extracted this band from the gel, and peptide MS/MS analysis confirmed that this protein was the product of the *fliC* gene. In addition, the majority of the peptides in the tryptic digest were readily identified by nLC-MS/MS, giving 62% sequence coverage. Tryptic peptides that had previously been observed to harbor the 398-Da glycan modification were observed to be unmodified on flagellin prepared from the *C. difficile* 630::0240*erm* strain. For example, the peptide T$^{202-212}$TMVSSLDAAK was observed as a doubly charged ion at *m/z* 730.8 in *C. difficile* 630. The same peptide was identified from the *C. difficile* 630::0240*erm* mutant as a doubly charged ion, at *m/z* 532.3, corresponding to the predicted mass of the unmodified peptide. The MS/MS spectrum of this peptide showed a clear sequence of peptide type *y* and *b* ions, corresponding to the amino acid sequence T$^{202-212}$TMVSSLDAALK. Notably, there were no glycan-associated ions observed in the low-mass region of the MS/MS spectrum, confirming that this peptide was unmodified in the absence of the glycosyltransferase (data not shown). Glycosylation appears to be required for optimal flagellin protein production, as the amounts produced by *C. difficile* 630::0240*erm* were much reduced, as evidenced by SDS-PAGE analysis. Negative staining of bacterial cells and examination by electron microscopy revealed that, as expected, no flagellar filaments were produced in the *fliC* mutant strain (Fig. 2D). In contrast, negative staining of *C. difficile* 630::0240*erm* cells revealed that flagellar filaments were produced but only in limited amounts compared to those of the parent strain, and these filaments appeared to be truncated in length (Fig. 2, compare panels A and C).

Motility screening of *fliC* and CD0240 mutant strains. Each strain was grown overnight on supplemented BHI agar containing Erm. A loopful of cells were used to stab inoculate motility agar tubes as described above. The *C. difficile* 630::*fliC*erm and *C. difficile* 630::0240*erm* strains no longer displayed a spreading phenotype in this motility agar, in contrast to the parent strain (Fig. 7B).

DISCUSSION

We demonstrate in this study that the flagellin proteins produced by *C. difficile* are glycoproteins with unique modifications. *C. difficile* 630 produces flagellin, which is glycosylated in O linkage at up to seven sites with a HexNAc residue, to which a methylated aspartic acid is linked via a phosphate bond. In contrast, flagellins from a number of *C. difficile* isolates from more recent outbreaks are modified in O linkage with a heterogeneous glycan containing up to five monosaccharide residues with masses of 204 (HexNAc), 146 (deoxyhexose), 160 (methylated deoxyhexose), and 192 (heptose). The compositions of the *C. difficile* glycans are quite distinct from that observed recently for flagellin from the related organism *C. botulinum*. The flagellar glycans characterized from *C. botulinum* isolates were shown to be either derivatives of the novel prokaryotic nonulosonate sugars or di-N-acetylhexuronic acids (7, 32, 33).

All *C. difficile* flagellins examined were shown to produce a glycan which is attached to serine and threonine residues in the protein sequence via a linking HexNAc monosaccharide. This observation suggested that the glycan biosynthesis assembly machinery in *C. difficile* is likely to be conserved, at least in part, among all strains. To explore this possibility, bioinformatic analysis of all *C. difficile* genomes identified a homolog of a gene in *C. difficile* 630 (CD0240) which encodes a glycosyltransferase enzyme. This gene lies immediately downstream of the *fliC* gene in all strains. In this study, we show that inactivation of CD0240 gene by using the ClosTron mutagenesis system results in cells which are no longer motile and can produce only limited amounts of unglycosylated flagellin. It remains to be determined if the product of the CD0240 gene is responsible for the transfer of only the initial HexNAc sugar to the flagellin protein or for the addition of the complete glycan,

but from this study, we show that the enzyme activity is required for glycosylation to occur. We also show that glycosylation of the flagellin protein is required for proper assembly and consequent motility.

This process of flagellar glycosylation has also been shown to be required for motility in a number of other bacterial pathogens, including *Campylobacter jejuni*, *Helicobacter pylori*, and *Aeromonas caviae*, which are all known to colonize the gastrointestinal tract. In contrast, O-linked flagellar glycosylation, which occurs in *Pseudomonas aeruginosa*, *Pseudomonas syringae*, and *Listeria monocytogenes*, is not required for flagellar assembly, although in the case of *P. aeruginosa* and *P. syringae* pv. tabaci 6605, the glycan has been implicated in virulence (2, 28). The current study provides the first example of a gram-positive anaerobe in which the glycosylation process is required for flagellar assembly. While the flagellin of *Clostridium botulinum* was shown to be glycosylated, the role of the glycosylation process in assembly and motility had not been investigated. With the recent development of mutagenesis tools for clostridial species, it will now be possible to determine if the glycosylation process is also required for flagellar assembly in *C. botulinum*. The precise biological role of the glycosylation process in flagellar assembly has yet to be determined. It may be required for stability of subunit-subunit interactions within the flagellar filament or alternatively required for efficient secretion of the flagellin monomer through the basal body apparatus.

As the flagellar glycans from the clinical isolates are quite distinct in structure from that made by *C. difficile* 630 and there appears to be conservation in the genetic content of this locus among the seven recent genome-sequenced isolates as well as the clinical isolates examined in the current study, it is possible that the glycan may also contribute to the pathogenic potential of these strains. Future work will be directed toward defining whether motility or the glycans play a role in colonization and in determining if there is a correlation between glycan structure and virulence of isolates.

It is important to acknowledge that two recent genomic microarray studies suggest that motility may not be required for virulence due to divergence in content of flagellar coding sequences among strains (14, 26). However, in the study by Stabler et al. (26), the group of genes comprising the flagellar glycosylation locus from QCD32g58 were not included in the microarray. In the study by Janvilisri et al. (14), primers of approximately 70 bp specific for each of the flagellar glycosylation island genes from the QCD32g58 locus were included in the microarray, in addition to primers specific for CD0240 from *C. difficile* 630. Of the 35 human strains tested by microarray, only 3 strains were missing a single gene from this locus, while 2 strains appeared to be missing three of the genes (Yung-Fu Chang, personal communication), supporting the observation in the current study that this locus is conserved among human isolates. In contrast to the results obtained by microarray, for which it appears that the CD0240 gene is divergent or absent in a significant number of isolates, we show by PCR that a homolog of CD0240 is present in all genome-sequenced isolates as well as in clinical isolates in the current study. It should be noted that sequence diversity in the region of this gene which spans one of the primers used in the microarray may be responsible for the results obtained in the analysis of

Janvilisri et al. (14), which indicated that the gene was divergent or absent in many isolates.

While motility may not be an essential virulence factor for all isolates, it remains to be established if other surface-associated proteins are also glycosylated with the products of this locus, thereby allowing the novel glycan structure to play a role in colonization during infection.

REFERENCES

1. **Alfa, M. J., T. Du, and G. Beda.** 1998. Survey of incidence of *Clostridium difficile* infection in Canadian hospitals and diagnostic approaches. J. Clin. Microbiol. **36:**2076–2080.
2. **Arora, S. K., A. N. Neely, B. Blair, S. Lory, and R. Ramphal.** 2005. Role of motility and flagellin glycosylation in the pathogenesis of *Pseudomonas aeruginosa* burn wound infections. Infect. Immun. **73:**4395–4398.
3. **Barken, K. B., S. J. Pamp, L. Yang, M. Gjermansen, J. J. Bertrand, M. Klausen, M. Givskov, C. B. Whitchurch, J. N. Engel, and T. Tolker-Nielsen.** 2008. Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in *Pseudomonas aeruginosa* biofilms. Environ. Microbiol. **10:**2331–2343.
4. **Bédouet, L., F. Arnold, G. Robreau, P. Batina, F. Talbot, and A. Binet.** 1998. Evidence for an heterogeneous glycosylation of the *Clostridium tyrobutyricum* ATCC 25755 flagellin. Microbios **94:**183–192.
5. **Borriello, S. P.** 1998. Pathogenesis of *Clostridium difficile* infection. J. Antimicrob. Chemother. **41**(Suppl. C):13–19.
6. **Calabi, E., S. Ward, B. Wren, T. Paxton, M. Panico, H. Morris, A. Dell, G. Dougan, and N. Fairweather.** 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. Mol. Microbiol. **40:**1187–1199.
7. **Carter, A. T., C. J. Paul, D. R. Mason, S. M. Twine, M. J. Alston, S. M. Logan, J. W. Austin, and M. W. Peck.** 2009. Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*. BMC Genomics **10:**115.
8. **Coon, J. J., J. Shabanowitz, D. F. Hunt, and J. E. Syka.** 2005. Electron transfer dissociation of peptide anions. J. Am. Soc. Mass Spectrom. **16:**880–882.
9. **Emerson, J. E., C. B. Reynolds, R. P. Fagan, H. A. Shaw, D. Goulding, and N. F. Fairweather.** 2009. A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. Mol. Microbiol. [Epub ahead of print.] doi:10.1111/j.1365-2958.2009.06812.x.
10. **Fekety, R., L. V. McFarland, C. M. Surawicz, R. N. Greenberg, G. W. Elmer, and M. E. Mulligan.** 1997. Recurrent *Clostridium difficile* diarrhea: characteristics of and risk factors for patients enrolled in a prospective, randomized, double-blinded trial. Clin. Infect. Dis. **24:**324–333.
11. **Guerry, P., C. P. Ewing, M. Schirm, M. Lorenzo, J. Kelly, D. Pattarini, G. Majam, P. Thibault, and S. Logan.** 2006. Changes in flagellin glycosylation affect Campylobacter autoagglutination and virulence. Mol. Microbiol. **60:**299–311.
12. **Heap, J. T., O. J. Pennington, S. T. Cartman, G. P. Carter, and N. P. Minton.** 2007. The ClosTron: a universal gene knock-out system for the genus Clostridium. J. Microbiol. Methods **70:**452–464.
13. **Hogan, J. M., S. J. Pitteri, P. A. Chrisman, and S. A. McLuckey.** 2005. Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. J. Proteome Res. **4:**628–632.
14. **Janvilisri, T., J. Scaria, A. D. Thompson, A. Nicholson, B. M. Limbago, L. G. Arroyo, J. G. Songer, Y. T. Grohn, and Y. F. Chang.** 2009. Microarray identification of *Clostridium difficile* core components and divergent regions associated with host origin. J. Bacteriol. **191:**3881–3891.
15. **Jiménez, C. R., S. Spijker, S. de Schipper, J. C. Lodder, C. K. Janse, W. P. Geraerts, J. van Minnen, N. I. Syed, A. L. Burlingame, A. B. Smit, and K. Li.** 2006. Peptidomics of a single identified neuron reveals diversity of multiple neuropeptides with convergent actions on cellular excitability. J. Neurosci. **26:**518–529.

16. **Kyne, L., R. J. Farrell, and C. P. Kelly.** 2001 *Clostridium difficile*. Gastroenterol. Clin. N. Am. **30:**753–777, ix–x.

17. **Logan, S. M.** 2006. Flagellar glycosylation—a new component of the motility repertoire? Microbiology **152:**1249–1262.

18. **Loo, V. G., L. Poirier, M. A. Miller, M. Oughton, M. D. Libman, S. Michaud, A. M. Bourgault, T. Nguyen, C. Frenette, M. Kelly, A. Vibien, P. Brassard, S. Fenn, K. Dewar, T. J. Hudson, R. Horn, P. Rene, Y. Monczak, and A. Dascal.** 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. N. Engl. J. Med. **353:**2442–2449.

19. **McDonald, L. C., G. E. Killgore, A. Thompson, R. C. Owens, Jr., S. V. Kazakova, S. P. Sambol, S. Johnson, and D. N. Gerding.** 2005. An epidemic, toxin gene-variant strain of *Clostridium difficile*. N. Engl. J. Med. **353:**2433–2441.

20. **Pépin, J., L. Valiquette, M. E. Alary, P. Villemure, A. Pelletier, K. Forget, K. Pepin, and D. Chouinard.** 2004. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. CMAJ **171:**466–472.

21. **Pépin, J., L. Valiquette, and B. Cossette.** 2005. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. CMAJ **173:**1037–1042.

22. **Samatey, F. A., K. Imada, S. Nagashima, F. Vonderviszt, T. Kumasaka, M. Yamamoto, and K. Namba.** 2001. Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. Nature **410:**331–337.

23. **Schirm, M., M. Kalmokoff, A. Aubry, P. Thibault, M. Sandoz, and S. M. Logan.** 2004. Flagellin from *Listeria monocytogenes* is glycosylated with beta-O-linked *N*-acetylglucosamine. J. Bacteriol. **186:**6721–6727.

24. **Schirm, M., I. C. Schoenhofen, S. M. Logan, K. C. Waldron, and P. Thibault.** 2005. Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. Anal. Chem. **77:**7774–7782.

25. **Sebaihia, M., B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R. Thomson, A. P. Roberts, A. M. Cerdeno-Tarraga, H. Wang, M. T. Holden, A. Wright, C. Churcher, M. A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S. Moule, K. Mungall, C. Price, E. Rabbinowitsch, S. Sharp, M. Simmonds, K. Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell, and J. Parkhill.** 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat. Genet. **38:**779–786.

26. **Stabler, R. A., D. N. Gerding, J. G. Songer, D. Drudy, J. S. Brazier, H. T. Trinh, A. A. Witney, J. Hinds, and B. W. Wren.** 2006. Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. J. Bacteriol. **188:**7297–7305.

27. **Syka, J. E., J. J. Coon, M. J. Schroeder, J. Shabanowitz, and D. F. Hunt.** 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. USA **101:**9528–9533.

28. **Taguchi, F., K. Takeuchi, E. Katoh, K. Murata, T. Suzuki, M. Marutani, T. Kawasaki, M. Eguchi, S. Katoh, H. Kaku, C. Yasuda, Y. Inagaki, K. Toyoda, T. Shiraishi, and Y. Ichinose.** 2006. Identification of glycosylation genes and glycosylated amino acids of flagellin in *Pseudomonas syringae pv. tabaci*. Cell. Microbiol. **8:**923–938.

29. **Tasteyre, A., M. C. Barc, A. Collignon, H. Boureau, and T. Karjalainen.** 2001. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. Infect. Immun. **69:**7937–7940.

30. **Tasteyre, A., M. C. Barc, T. Karjalainen, P. Dodson, S. Hyde, P. Bourlioux, and P. Borriello.** 2000. A *Clostridium difficile* gene encoding flagellin. Microbiology **146**(Part 4):957–966.

31. **Tasteyre, A., T. Karjalainen, V. Avesani, M. Delmee, A. Collignon, P. Bourlioux, and M. C. Barc.** 2000. Phenotypic and genotypic diversity of the flagellin gene (*fliC*) among *Clostridium difficile* isolates from different serogroups. J. Clin. Microbiol. **38:**3179–3186.

32. **Twine, S. M., C. J. Paul, E. Vinogradov, D. J. McNally, J. R. Brisson, J. A. Mullen, D. R. McMullin, H. C. Jarrell, J. W. Austin, J. F. Kelly, and S. M. Logan.** 2008. Flagellar glycosylation in *Clostridium botulinum*. FEBS J. **275:**4428–4444.

33. **Voisin, S., R. S. Houliston, J. Kelly, J. R. Brisson, D. Watson, S. L. Bardy, K. F. Jarrell, and S. M. Logan.** 2005. Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. J. Biol. Chem. **280:**16586–16593.

34. **Voth, D. E., and J. D. Ballard.** 2005. *Clostridium difficile* toxins: mechanism of action and role in disease. Clin. Microbiol. Rev. **18:**247–263.

35. **Wilcox, M. H., and R. C. Spencer.** 1992. *Clostridium difficile* infection: responses, relapses and re-infections. J. Hosp. Infect. **22:**85–92.

36. **Wilkins, T. D., and D. M. Lyerly.** 2003. *Clostridium difficile* testing: after 20 years, still challenging. J. Clin. Microbiol. **41:**531–534.

37. **Yildiz, F. H., and K. L. Visick.** 2009. Vibrio biofilms: so much the same yet so different. Trends Microbiol. **17:**109–118.