NRC Publications Archive
Archives des publications du CNRC

The type B flagellin of hypervirulent clostridium difficile is modified with novel sulphonated Peptidylamido-glycans
Bouché, Laura; Panico, Maria; Hitchen, Paul; Binet, Daniel; Sastre, Federico; Faulds-pain, Alexandra; Valiente, Esmeralda; Vinogradov, Evgeny; Aubry, Annie; Fulton, Kelly; Twine, Susan; Logan, Susan m; Wren, Brendan w; Dell, Anne; Morris, Howard r

National Research Council Canada
Conseil national de recherches Canada

Clostridium difficile type B flagellin is glycosylated with sulphonated peptidylamido-glycans

The Type B Flagellin of Hypervirulent *Clostridium difficile* is Modified with Novel Sulphonated Peptidylamido-Glycans

Laura Bouché[1], Maria Panico[1], Paul Hitchen[1], Daniel Binet[4], Federico Sastre[1#], Alexandra Faulds-Pain[2], Esmeralda Valiente[2], Evgeny Vinogradov[3], Annie Aubry[3], Kelly Fulton[3], Susan Twine[3], Susan M. Logan[3], Brendan W. Wren[2], Anne Dell[1*] and Howard R. Morris[1,4]

[1] Department of Life Sciences, Imperial College London, South Kensington Campus, SW7 2AZ, London, UK.

[2] Department of Pathogen Molecular Biology. London School of Hygiene and Tropical Medicine. Keppel Street, WC1E 7HT, London, UK.

[3] Vaccine Program, Human Health Therapeutics Portfolio, National Research Council, Ottawa, ON K1A 0R6, Canada.

[4] BioPharmaSpec, Suite 3.1 Lido Medical Centre, St. Saviours Road, Jersey, JE2 7LA, GB.

[#] Current address: Departamento de Ingeniería Química y Bioprocesos, Escuela de Ingeniería, Pontificia Universidad Católica de Chile (PUC). Av. Vicuña Mackenna 4860, Santiago de Chile, 7820436, Chile

Running title: *Clostridium difficile* type B flagellin is glycosylated with sulphonated peptidylamido-glycans

*Corresponding author information:
email: a.dell@imperial.ac.uk
Imperial College London,
Department of Life Sciences,
South Kensington Campus,
London, SW7 2AZ, UK
Phone: +44 (0) 2075945219
Fax: + 44 (0) 207 2250458

Abstract

Glycosylation of flagellins is a well recognized property of many bacterial species. In this study we describe the structural characterization of novel flagellar glycans from a number of hypervirulent strains of *C. difficile*. We used mass spectrometry (nano LC- MS and MS/MS analysis) to identify a number of putative glycopeptides which carried a variety of glycoform substitutions each of which was linked through an initial HexNAc residue to Ser or Thr. Detailed analysis of a LLDGSSTEIR glycopeptide released by tryptic digestion, which carried two variant structures, revealed that the glycopeptide contained, in addition to carbohydrate moieties, a novel structural entity. A variety of Electrospray-MS strategies using Q-TOF technology were used to define this entity, including positive- and negative-ion collisionally activated decomposition (CAD) MS/MS which produced unique fragmentation patterns, and high resolution accurate mass measurement to allow derivation of atomic compositions, leading to the suggestion of a Taurine-containing peptidylamido-glycan structure. Finally NMR analysis of flagellin glycopeptides provided complementary information. The glycan portion of the modification was assigned as α-Fuc3N-(1→3)-α-Rha-(1→2)-α-Rha3OMe-(1→3)-β-GlcNAc-(1→)Ser and the novel capping moiety was shown

1

to be comprised of Taurine, Alanine, and Glycine. This is the first report of a novel O-linked sulphonated peptidylamido-glycan moiety decorating a flagellin protein.

Introduction

The intestinal pathogen *Clostridium difficile* is the leading cause of antibiotic-associated diarrhea worldwide. The pathogen colonizes the gastro-intestinal tract when the normal microbiota is disturbed after antibiotic treatment, causing *C. difficile* infection (CDI) in susceptible patients. In the past decade, CDI mortality has increased dramatically since the emergence of hypervirulent strains such as the PCR ribotype (RT) 027 and 023 lineages (1,2).

The reasons for the increase in severity of CDI caused by hypervirulent strains are still not well understood. Molecules implicated in *C. difficile* virulence include the secreted toxins TcdA and TcdB, and a variety of cell surface biopolymers (3-6). Amongst the latter, flagella, which are responsible for the pathogen's motility, are believed to have roles in virulence because disruption of their biosynthesis and expression affects colonisation, biofilm formation and toxin production (7).

The flagellin proteins of *C. difficile* are known to be post-translationally modified with *O*-linked glycans and there is evidence that glycosylation can affect motility and virulence (8,9). Flagellin *O*-glycosylation is widespread in Gram negative bacteria (10) but so far has only been found in three Gram positive genera, *Clostridium*, *Listeria* and *Paenibacillus* (11-13). A great deal of diversity exists amongst flagellin glycans but there are some common themes. For instance, many Gram negative flagellins have a single pseudaminic acid or legionaminic acid residue at each of their *O*-glycosylation sites. These sugars can be substituted with a variety of functionalities such as acyl and acetamido groups resulting in considerable structural heterogeneity. Interestingly, the first *Clostridium* species to have its flagellin glycosylation characterised, *Clostridium botulinum,* was found to share this type of glycosylation. Thus, its flagellin is substituted with the legionaminic acid derivative 7-acetamido-5-(N-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-glycero-alpha-D-galacto-nonulosonic acid (alphaLeg5GluNMe7Ac)(12) In contrast, *C. difficile* post-translational modifications appear to be quite different in structural composition from those which have been found in Gram negative organisms.

The best characterized *C. difficile* flagellin is that from the first strain to have its genome sequenced (14). This PCR-ribotype 012 strain (630) was isolated from a Swiss hospital patient in 1982. It is an epidemic, multidrug-resistant strain and pre-dates the emergence of the hypervirulent strains. The 630 flagellin is modified at up to seven sites with the monosaccharide N-AcetylGlucosamine (GlcNAc) which is substituted with a phosphodiester-linked N-methyl-L-Threonine residue (8,9). RT027 and RT023 strains, however, appear to lack this unusual amino acid modification. Preliminary data from mass spectrometric (MS) investigations of RT027 flagellins was interpreted as consistent with the presence of HexNAc-linked oligosaccharides up to pentasaccharide in length. Mass increments in the MS data were attributed to various compositions of deoxyhexose, methylated deoxyhexose, HexNAc and heptose, but full structures were not defined at that time because of technological limitations (8).

In this study, we have rigorously characterized flagellin glycosylation in several emerging hypervirulent clonal strains including RT027, RT023, RT106 and RT001. We have discovered several variants of a novel peptidylamido-glycan sulphonate structure decorating hydroxyamino acids (Ser and Thr) in the flagellin (FliC) protein. Here we describe the chemical characterization of this newly discovered structure using Q-TOF technology (15,16) applied in a variety of structure elucidation strategies (17,18) including positive- and negative-ion collisionally activated decomposition (CAD) MS/MS to produce unique and interpretable fragmentation patterns, and high resolution accurate mass measurement to allow derivation of detailed atomic compositions, then confirming and extending the structural interpretations of the MS data by purifying a number of glycopeptides in sufficient quantities to define each structure by NMR.

Results

Mass Spectrometric Analysis of Flagellin Glycopeptides

A preliminary structure of RT027 flagellin

2

glycosylation had been previously studied by LC-MS/MS analysis (8) but full structures were not solved due to technical limitations. In the study presented here, following nano LC-MS and MS/MS analysis, a number of putative glycopeptides derived from RT0207 and other flagellins were identified via a search for oxonium ions (eg m/z 204) in the data sets. This led to the discovery of a number of glycopeptide candidates carrying a variety of glycoform substitutions including the peptides LLDGSSTEIR, VALVNTSSIMSK and QMVSSLDVALK. Each of these peptides carried a glycan modification which was linked through an initial HexNAc residue to Ser or Thr followed by deoxyHex/MethyldeoxyHex and MethyldeoxyHex/deoxyHex and MethyldeoxyHex/MethyldeoxyHex in positions 2 and 3 respectively. The majority of structural work reported here has concentrated on the LLDGSSTEIR glycopeptides from strain R20291, and interestingly, two variant structures carried on this peptide were present in the LCMS chromatograms at m/z 998^{2+} and 991^{2+} where the MS/MS fragmentation pattern clearly showed (in addition to carbohydrate moieties) the presence of a novel structural entity not previously observed in sugar or amino acid chemistry.

The collisionally activated decomposition (CAD) MS/MS spectra of m/z 998^{2+} and 991^{2+} molecules are shown in **Figure 1**, with the main spectrum showing the high resolution data (to 4 decimal places) for the larger molecule (m/z 998^{2+}), and inset showing the low resolution spectrum of m/z 991^{2+} for comparison. From these data, it is clear that a HexNAc-methyldeoxyHex-deoxyHex-glycosyl substituent is attached to the peptide backbone (seen at 1090, M+H$^+$) via signals at m/z 1293, 1453, and 1599 respectively (corresponding to glycosidic cleavages), together with a series of y'' ions beginning at m/z 749 and extending with the same substituents, identically for both precursor ions. A further less intense signal is present at m/z 1744 in the high-resolution data extending the glycosylation sequence by an amino-dideoxyHex unit. The high resolution mass measurement of the 991^{2+} signal compared with the 998^{2+} signal shows that the 14 amu mass difference corresponds to a CH$_2$ difference between the two structures. The clearly novel aspect of these glycopeptides can be seen in the

substantial fragment ions at m/z 396, 378, 268, 251, 223 and 152 which do not immediately correlate with sugar or amino acid origin. The equivalent signals are present in the 991^{2+} data (inset spectrum) 14 Da lower at m/z 382, 364, 254, 237, 209 except for m/z 152 (same), showing that the 14 Da mass difference resides between the 152 and 396/382 fragments in the structure. When cross-correlating the observed 396/382 and 1599 signals with the molecular masses observed in the 998^{2+}/991^{2+} quasimolecular ions then these fragments are additive to the molecular mass (allowing for hydrogen transfers) and thus represent between them the overall glycopeptide structure, as shown schematically in **Figure 1** for the 998^{2+} variant.

The interpretation of the mechanisms leading to these fragment ions was greatly assisted by the atomic compositions determined from the accurate masses in the high resolution Q-TOF data (see representative data in **Table 1**) and also by the presence of "counterion" data, as is often observed in doubly charged MS/MS spectra. For example, the aminodideoxyHex residue extension from m/z1599 to 1744 must therefore be present in the m/z 396 counterpart, and its partial loss (128Da) by beta-elimination is observed to give m/z 268 in **Figure 1**, whereby the amino function is retained by the 268 ion, which itself then loses first ammonia to m/z 251and then carbon monoxide to m/z 223 from the amide group thus assigned. The charge can of course be retained for a proportion of fragments on the "counterion" instead of m/z 268 which would be present at m/z 129 from the elimination mechanism. The significant signal at m/z 111 (C$_6$H$_7$O$_2$) was assigned to that origin via loss of water to give a highly stable triply conjugated cyclic ion, available without rearrangement by postulating a 3-amino substitution. The next significant ion in the low mass region is seen at m/z 152 and it was recognised that the observed accurate mass difference of 71.037 Da (223.0753 minus 152.0383), could correspond to either an Alanine or its isomer N-methyl Glycine (atomic composition C$_3$H$_5$NO, theoretical mass 71.0371). In mechanistic terms, the carbonyl of such a unit could be forming the amide linkage to the amino sugar in m/z 396 which would predictably fragment to give the 268, 251 (b$_2$) and 223 (a$_2$) ions and then to give a terminal "b$_1$" ion at m/z

3

180 (from loss of Ala or N(Me)Gly) the signal for which is not observed. However, the m/z 152 mass is 28 Daltons (CO from the atomic composition) below this mass and it is very common for the a_1 ion subfragment (an aldimine) to be the more intense ion in peptide fragmentation, particularly if the Nitrogen is alkylated by methyl or another grouping giving rise to a tertiary Nitrogen. The summary of the interpretation logic used to assign the above MS signals at that stage of the study is shown in **Scheme 1**.

The m/z 152 signal was found to possess an unusually mass-deficient accurate mass which suggested a sulphur-containing atomic composition determined as $C_4H_{10}NO_3S$ for this terminal fragment, which is not the formula of any previously reported protein- or carbohydrate-derived structural unit. In order to help define this fragment in more detail, and to provide supporting evidence for the ideas in **Scheme 1** two further sets of experiments were carried out on the remaining small quantities of material: (a) MS/MS analysis of several of the key fragment ions above to confirm mechanistically understandable breakdown products and (b) experiments in the negative ion MS and MS/MS modes to look for new and complementary fragment ion information. CAD MS/MS of the cone voltage induced m/z 152 from the 991^{2+} and 998^{2+} glycopeptide samples gives rise to two principal ion species seen in **Figure 2** at m/z 108 ($C_2H_6NO_2S$) and m/z 70 (C_4H_8N) suggesting two overlapping component fragments (a sulphonic acid and an alkylamine) competitively derived from the m/z 152 ion ($C_4H_{10}NO_3S$). The negative ion CAD MS/MS spectrum obtained for m/z 996^{2-} is shown in **Figure 3**. Firstly, these data confirm the basic structural features of the novel glycosylation inferred from the positive ion data in **Figure 1** and shown in **Scheme 1** regarding expected principal fragments at m/z 921, 718, 540, 394, 266, 180 and 150, with m/z 718 for example corresponding to X-aminodideoxyHex-deoxyHex-methyldeoxyHex⁻. Some other signals are derived from the carrier peptide LLDGSSTEIR but, importantly, a new fragment ion, (with the equivalent not being observed in the positive ion MS/MS), is seen at m/z 124, and this was recognised as corresponding to the mass of aminoethyl-sulphonic acid, Taurine ($C_2H_6NO_3S$, M-H⁻). The MS/MS of this 996^{2-}-derived signal gave rise to an identical MS/MS

spectrum (**Figure 4)** to that observed for a synthetic sample of Taurine itself, being mainly a SO3⁻ fragment ion base peak, thus providing strong evidence for this structural unit within the m/z 152 ion in the novel structure.

Further experiments involving subfragment MS/MS, and also hydrogen/deuterium exchange to count heteroatom-linked protons, were carried out to validate the structural conclusions from the mass spectrometric experiments (data not shown), but as seen in the structural summary in **Scheme 1** an ambiguity nevertheless remains in the proposed novel m/z 268 unit whereby although the amide-linked aminoacid to the aminodideoxyHex would be Glycine in the 991^{2+} structure, it could be either Alanine or N-methyl Glycine in the 998^{2+} structure. There are also several possible ways of arranging the R, R_1 and R_2 atoms in the unit incorporating the aminoethyl sulphonic acid-containing 152 (180) structure, including variants containing Taurine itself or Cysteic acid with alkylations, to produce the necessary accurate masses observed thus satisfying the atomic compositions determined from high resolution mass measurement. To fully characterize the R-groups in **Scheme 1** and to define the stereochemistry and linkages, a scaled-up preparation of flagellin was made which was used to then allow the isolation of sufficient material for the derivation of NMR data.

NMR Analysis of flagellin glycopeptides

Flagellin protein (15mg) was extensively digested with Proteinase K to enable purification of sufficient quantities of glycan material with minimal peptide backbone for NMR structural analysis. This proteinase K digested material was subjected to fractionation to obtain glycan enriched material using size exclusion (Biogel P10) chromatography followed by Zorbax C18 reverse phase. All fractions were analysed by ¹H NMR to identify fractions containing glycopeptide material. Three fractions were obtained following chromatographic separation which contained sufficient amounts of glycopeptide for further structural characterisation. Fraction 21 from the initial Zorbax C18 column separation contained a mixture of two glycan species (compounds 1 and 2, **Figure 5**). Fractions 23 and the re-separated fraction 12 contained compounds 2 and 3 (**Fig 5**).

2D NMR spectra (COSY, TOCSY, ROESY, ¹H-¹³C HSQC, HMBC, and HSQC-

4

TOCSY were recorded for all glycopeptide fractions. Spectra of fraction 21 (**Table 2**) contained spin-systems of α-Rha3OMe (A, B), β-GlcNAc (C), two Serine and one Threonine residues. The position of the methyl group on Rha3OMe was determined from HMBC correlation between Me and H/C-3 of this sugar. The sequence of the monosaccharides followed from NOE and HMBC correlations between H-1 of residue B to H/C-2 of the residue A, and between H-1 of A to H/C-3 of β-GlcNAc C. H-1 of GlcNAc C showed NOE correlation to Ser* H-3 and HMBC correlation to Ser* C-3, indicating that β-GlcNAc was linked to O-3 of a Serine residue (Ser*). Thus the oligosaccharide 1 had the structure shown in Fig. 5. The purified glycopeptide contained two Serine residues and one Threonine residue.

Another glycopeptide fraction obtained after re-separation contained a mixture of compounds 2 and 3 (**Figure. 5**). Compound 2 contained all components of the oligosaccharide 1 without an O-Me-group on O-3 of the terminal Rha, and additionally a residue of 3-amino-3,6-dideoxy-α-galactopyranose (Fuc3N), acylated at the aminogroup with the unusual sulphopeptide. The position of Fuc3N at O-3 of the Rha E followed from NOE and HMBC correlations F1:E3. The glycopeptide contained three components: Taurine, Alanine, and Glycine. The sequence of these components was determined based on the following observations: Glycine C-2 was observed at its usual place around 44ppm providing clear evidence that it was acylated by the Alanine moiety. In contrast, the low field position of the Ala C-2 at 57.5 ppm indicated that its NH group is alkylated, (rather than if it was within a peptide where the normal field position would be ~ 51 ppm). This provides evidence that the Taurine unit (deduced from the independent mass spectrometric analysis) is linked to the Alanine, and that Glycine would be acylating N-3 of Fuc3N, supporting the interpretation of the mass spectrometric data. The sequence of the sulphopeptide would then be Taurinyl-Ala-Gly as indicated in **Figure 5** (Compound 2).

Another component of this fraction was compound 3, which had the same sugar structure as compound 1 but was linked to a Threonine residue. The ^{1}H and ^{13}C NMR data for the glycopeptides present in fraction 21 is shown in **Table 2** and **Table 3** and the HSQC spectrum shown in **Figure 6.**

This NMR data show that the R_2 group in **Scheme 1** of the mass spectrometric interpretation is a methyl group (thus allowing an Alanine assignment) to give an Ala-Gly dipeptide, with R being a hydrogen and R_1 the ethyl sulphonic acid group giving rise to Taurine in the MS/MS fragmentation. The ^{13}C chemical shifts observed were in good agreement for an alkylated NH-group in the Alanine unit and with the shifts reported for Taurine [http://www.bmrb.wisc.edu/metabolomics/mol_su mmary/show_data.php?molName=taurine&id=bm se000120&whichTab=1]. The glycan portion of the structure, assigned mass spectrometrically as an amino-dideoxyHex-deoxyHex-methyldeoxyHex-HexNAc- unit in **Figure 1** was then further assigned from the NMR data as α-Fuc3N-(1→3)-α-Rha-(1→2)-α-Rha3OMe-(1→3)-β-GlcNAc-(1→)Ser. The overall structure of the "991" peptidyl-glycan is shown as compound 2 in **Figure 5**, and our mass spectrometric data suggests that the corresponding "998^{2+}" peptidyl-glycan will have either Alanine or N-methyl Glycine replacing the Glycine in the dipeptide portion of that structure producing the observed 14 mass unit shift. Although the NMR spectra of compounds 2 and 3 contained additionally a minor methyl group signal, at 3.21/55.1 ppm, which may be responsible for the observed mass increase of 14 amu in the mass spectrum of some glycopeptides ("998"), we were unable to assign the position of this methyl group within the structure.

Discussion

Glycosylation is a key modification of proteins and lipids that is often important in intermolecular and intercellular interactions. Bacterial protein glycosylation systems have come under enhanced scrutiny because of the increasing association with pathogenic species. Recent research is providing compelling evidence for protein glycosylation being central to the survival and pathogenesis of many bacteria. They have been described variously as being important in adhesion, motility, DNA uptake, biofilm formation, autoaggregation, invasion, serum resistance, immune evasion and animal colonization (19-25) (1-7). Recently, the structure and biological role of flagellar

5

glycosylation in the enteric opportunistic pathogen *C.difficile* 630 has been described (8,9). Moreover, investigations into the biological role of flagellar glycosylation in the emerging hypervirulent *C.difficile* RT027 and RT023 have been undertaken and reported in our accompanying paper(26). In both *C.difficile* studies, flagella post-translational modification plays a role in motility, aggregation and adhesion to abiotic surfaces. In the case of *C.difficile* RT027 flagella glycosylation is also involved in Caco-2 cell adhesion. In the present study, we have revealed the discovery of a unique flagellar non-reducing end peptidylamido-glycan structure on glycoproteins isolated from RT027 and others (RT023, 001 and 106).

Following the mass spectrometric discovery of unusual post-translational modifications (PTMs) in the LC-MS and MS/MS data from tryptic digests of the FliC protein using an instrument tuned for optimal automated information-dependent acquisition, a full battery of advanced MS techniques (15-18) was first applied to characterise the novel components to the extent possible on the small quantities of protein available. These included the production of on-line high-resolution MS/MS data using a 40,000 resolving power Q-TOF geometry instrument (see Experimental Procedures) allowing the assignment of probable atomic compositions of all fragment ions, including those clearly not derived from normal glycopeptides or previously reported PTMs. These experiments showed for the first time the presence of sulphur-containing moieties, and subsequent MS^2 and MS^3 data generated by the high sensitivity and mass accuracy of the Xevo Q-TOF geometry instrument (15,16) in both positive and negative ion mode allowed the discovery of a Taurine (aminoethyl-sulphonic acid) unit in the breakdown fragments of several of the precursor ion species investigated. Combining all of these data sets, together with confirmatory experiments such as Hydrogen/Deuterium exchange analysis (not shown), allowed the novel structural unit containing the sulphonic acid group to be defined as shown in **Scheme 1** (with composition $C_4H_{10}NO_3S$). When it then became possible to isolate sufficient of the des-methyl (991[2+]) variant of the structure, a detailed NMR study was used to reveal the identity of the R, R_1 and R_2 groups shown to give the overall PTM structure in **Figure 5**. Our mass spectrometric

studies on hypervirulent strains Ribotype 027, 023, 106 and 001 have shown the presence of one of these structures ("991[2+]" or "998[2+]") and related variants (data not shown) in each of the strains studied. The work reported here has concentrated on what was found to be the most abundant LLDGSSTEIR tryptic glycopeptide, but preliminary mass spectrometric analysis shows that similar modifications are present on at least two other flagellin peptides, QMVSSLDVALK and VALVNTSSIMSK. Only minor amounts of the respective free (non-glycosylated) peptides were found in the digests.

In contrast to this very complex glycosylation of the flagella of *C.difficile* RT027, the 630 type strain is modified with single GlcNAc residues that are substituted with an N-methylated Threonine linked via a phosphodiester bond (8,9). Despite the substantial differences in glycosylation, a common feature is the presence of a negatively charged functionality in the periphery of the post-translational modification, namely a sulphonate in the hypervirulent strains and a phosphoester in the 630 strain. These charged groups are likely to be involved in ionic interactions between the flagella and extracellular structures. This could explain the phenotype of *C. difficile* flagellar glycosylation knockouts where autoaggregation, biofilm formation and adhesion to Caco-2 cells are reduced (9,26).

An increasing number of reports on flagellar glycosylation on Gram-positive and Gram-negative bacterial pathogens have been published (10,27). In comparison with Gram-negative bacteria, the reports on flagellar glycosylation on Gram-positive are limited. Among Gram-positive, Clostridium spp. are the most characterised (8,9,28-30). There are two other genera of Gram-positive pathogens with *O*-glycosylated flagella: Listeria and Paenibacilla. *Listeria monocytogenes* is glycosylated at up to six sites per monomer with a single β-*O*-linked GlcNAc residue (11). Paenibacilla has a flagella modified with an *O*-linked trisaccharide comprised of one hexose and two-N-acetyl-hexosamine residues, at three sites of glycosylation(13). In the case of Gram-negative bacteria, the diversity of flagellar glycosylation moieties is remarkable. Flagellins of many Gram-negative bacterial pathogens have a single pseudaminic or legionaminic acid residue at each of their *O*-glycosylation sites(10). Both these

6

sugars exhibit considerable diversity due to differences in acyl functionalities. Notably some are acylated with amino acids. Thus, Glycine and N-Acetyl Glutamine have each been observed as ester-linked substituents of pseudaminic acid in the flagella of *Aeromonas caviae* and *Campylobacter jejuni*, respectively (31). However there are no reports of peptidyl substituents on glycans from gram negative organisms. In addition, amino acid substituents have not previously been observed on the glycans of gram positive pathogens. Therefore our discovery that the *C.difficile* RT027 flagella are modified by a peptidylamido-sugar moiety is a unique finding within both Gram-positive and Gram-negative bacteria.

Another structural component, identified in the present study as a Taurine-like non-reducing end unit, is unique in flagella glycosylation from bacteria. This unit might be used by RT027 strains as a strategy to evade the host immune system as Taurine is reported to have a key role in the regulation of the innate immune response (32).

In some bacterial pathogens such as the enteric pathogen *Campylobacter jejuni*, glycosyltransferases and glycan biosynthetic genes are situated adjacent to biosynthetic flagellar genes (33). For *C.difficile* type B flagellin it appears that rhamnose biosynthetic genes (CDR20291_0223-0226) which are similar to *rmlD, rmlA, rmlC* and *rmlB*, lie upstream of *fliC* and are likely involved in the biosynthesis of the rhamnose moieties modifying the *C.difficile* RT027 flagella (26). Within the locus immediately downstream of *fliC* in addition to the three putative flagellar GT genes, there are a number of biosynthetic genes (CDR20291_0244-0247) which appear to be responsible for synthesis of the novel terminal moiety (26). These biosynthetic genes are present in the genomes of both RT027 and RT023 strains. The amino acid similarity of CDR20291_0247 to FdtB from *A. thermoaerophilus* suggests that this gene product is likely involved in the production of the FucN monosaccharide. Previous work using *Aneurinibacillus thermoaerophilus* and *Xamthomonas. campestris* enzymes have shown that RmlA and RmlB catalyze the first two steps for 3-amido-3,6-dideoxygalactose (FucN) biosynthesis(34) by producing the substrate dTDP-6-deoxy-D-xylohex-4-ulose. Optimally, this substrate would be converted to dTDP-6-deoxy-D-xylohex-3-ulose by FdtA, however it has been demonstrated that incubation of the 4-keto product with FdtB and cofactors will yield a moderate level of conversion to dTDP-Fucp3N in the absence of FdtA isomerase (34). It has been suggested that this may be due to production of 3-keto substrate via a non enzymatic processes (35). As no homolog of FdtA appears to be present in the *C. difficile* R20291 type B post-translational modification locus it appears that this may be the mechanism whereby limited amounts of FucN can be produced in *C. difficile* and incorporated into the flagellar glycan. The heterogeneity observed in flagellin glycan composition (sulphonated peptidylamido-glycan structure and truncated methylated trisaccharide (Rha-Rha-GlcNAc) structure) might be explained by the inefficient production of FucN from dTDP-6-deoxy-D-xylohexose-4-ulose by FdtB. A limiting amounts of FucN would prevent the synthesis of the full flagellar glycan. In addition to the FdtB homolog there exists a putative acyltransferase (CDR20291_0244) which is a candidate for acylating the amino group of the fucosamine moiety. Downstream of CDR20291_0244 lie CDR20291_0245, CDR20291_0246 which are co-transcribed with the GT genes and encode a putative D-alanine ligase and alanine dehydrogenase, respectively. These genes are likely involved in the biosynthesis of the peptidyl moiety of this novel flagellar glycan (26).

This study reveals a unique flagellar glycosylation structure in the bacterial pathogen *C. difficile* hypervirulent RT027 strains which could provide the organism with a novel strategy to escape the immune system and be more virulent. Furthermore, this work highlights the diversity of glycans modifying flagella in other hypervirulent *C.difficile* RT such as RT023, RT001 and RT106 which could suggest different strategies of *C.difficile* to evade the immune system.

.

Experimental Procedures

Bacterial growth and Cultures - *C. difficile* strains, used in this study, are shown in **Table 4**. Strains were routinely grown on Brazier's CCEY agar (BioConnection, Leeds, United Kingdom) containing 4% [wt/vol] egg yolk, 250 mg/ml cycloserine, 8 mg/ml cefoxitin (Bioconnections, Leeds, UK) and 1% defibrinated horse blood (TCS Biosciences, Buckingham, UK), in blood agar

7

(Oxoid, Hampshire, UK) and in BHI (brain heart infusion media [Oxoid, Hampshire, UK]) and BHIS (BHI supplemented with 0.5% w/v yeast [Sigma, Gillingham,UK], 0.1% L-Cysteine [Sigma, Gillingham, UK]) broth or agar. All cultures were incubated at 37 °C overnight in a Don Whitney MG500 anaerobic workstation (Don Whitney Scientific Ltd).

Protein Isolation and Digestion - Bacteria were grown on BHIS overnight, at 37 °C. *C. difficile* strains were harvested, washed in phosphate buffered saline and resuspended in a 1:100 volume of low-pH glycine (0.2 M glycine-HCl, pH 2.2) and incubated at room temperature for 30 min with gentle shaking. The cells were removed by centrifugation at 4 °C and the supernatant neutralized with the addition of 2 M Tris to a pH of 7 to 8. Flagellin preparations were analysed by 12% SDS-PAGE Novex® NuPAGE® tris-glycine SDS-PAGE. The separated proteins were stained with coomassie in accordance with standard techniques. Flagellin bands were excised lyophilized and digested with trypsin (E.C.3.4.21.4, Promega) overnight. Peptides were extracted from gel pieces and studied by-LC-MS/MS analysis.

Mass Spectrometry - Positive Ion Q-TOF technology (15,16) was used to study the tryptic digests of proteins isolated from the various strains of *C. difficile* ribotypes in LC-MS and MS/MS experiments conducted on several instruments. A Sciex QStar Pulsar 1 instrument with LC Packings nanoLC using a C-18 nanocapillary (75 micron x 15 cm Pepmap) column eluting with a 0.05% formic acid/5%-95% acetonitrile gradient over 90 min, and information dependent acquisition (IDA) data acquisition was used for the low resolution (5000 resolving power) analysis of digest mixtures. A Waters XevoG2 instrument with Acquity microbore UPLC using a C-18 (1mm x 50mm BEH) column eluting with a 0.1% formic acid/0%-85% acetonitrile gradient over 40 min, with data dependent acquisition (DDA) acquisition, was used for detailed positive- and negative-ion MS^n and cold isotope labelling experiments, and a Waters Synapt G2-S, with nanoAcquity UPLC and a 75 micron x 15 cm BEH column and gradient elution with 0.05% formic acid/5%-95% acetonitrile over 45 min, and DDA data acquisition, was used for the collection of the High Resolution (40,000 resolving power) MS/MS

data from which atomic compositions could be calculated. MS^3 data was generated on the XevoG2 instrument on various fragment ions of interest by nanospray of 1-2 microlitres of LC-collected fractions loaded into borosilicate needles, using high cone voltage to achieve source fragmentation, then passing the fragment ions of interest into the collision cell for CAD MS/MS.

Negative Ion mass spectrometry was used to study collected fractions of interest. These were adjusted to basic pH using 10% ammonia, then MS^3 data was generated on the XevoG2 instrument on various fragment ions of interest by nanospray as described above.

NMR glycan structural analysis - Large scale preparation of flagellin was made by growing *C. difficile* R20291 overnight on 40 BHIS agar plates at 37 °C in an anaerobic chamber. Cells were harvested into sterile distilled water (dH$_2$O) and cells were vortexed for 3 min to release flagella filaments from cells. Bacterial cells were removed by centrifugation at 13,000 RPM in an Eppendorf centrifuge for 5 min (x2). The supernatant was collected, pooled and passed through a 13mm PVDF filter. This sterile supernatant was then centrifuged at 55,000 rpm for 1 h at 4 °C in a Beckman Optima TLX Ultracentrifuge using a Beckman TLA 120.2 rotor. The pelleted material was resuspended in dH$_2$O and the ultracentrifugation step was repeated. The supernatant was discarded and the pellet resuspended in dH$_2$O. Flagellin protein was obtained from multiple bacterial growths (>10) to give a total of 10 mg flagellin protein at a concentration of 0.4mg/ml (as determined by Biorad Protein Assay). To obtain glycan material devoid of protein backbone for structural analysis, flagellin (10 mg) was digested with proteinase K at a ratio of 1:1 (Sigma) in 10mM Na$_2$PO$_4$ pH 7.6 at 37 °C for 48 h. The proteinase K digested material was lyophilized, resuspended in dH$_2$O and the sample fractionated by gel filtration on a Biogel P10 column (2.5x80 cm, 1% Acetic acid, RI detector). Each fraction was analyzed by ^{1}H NMR and the glycopeptide-containing fraction was then applied to a Zorbax C18 column in 0.1% TFA-80% acetonitrile gradient with UV detector at 220 nm and fractions collected and reexamined by ^{1}H NMR for presence of glycan. Separation was then repeated with selected fractions using a gradient

8

from 4 to 20% of 80% acetonitrile over 1 h. All fractions were then analyzed by NMR.

NMR spectroscopy: NMR experiments were carried out on a Bruker AVANCE III 600 MHz (^{1}H) spectrometer with 5 mm Z-gradient probe with acetone internal reference (2.225 ppm for ^{1}H and 31.45 ppm for ^{13}C) using standard pulse sequences cosygpprqf (gCOSY), mlevphpr (TOCSY, mixing time 120 ms), roesyphpr (ROESY, mixing time 500 ms), hsqcedetgp (HSQC), hsqcetgpml (HSQC-TOCSY, 80 ms TOCSY delay) and hmbcgplpndqf (HMBC, 100 ms long range transfer delay). Resolution was kept <3 Hz/pt in F2 in proton-proton correlations and <5 Hz/pt in F2 of H-C correlations. The spectra were processed and analyzed using the Bruker Topspin 2.1 program.

Conflict of interest- The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions

E.V.[2], A.D., B.W.W., and H.R.M. designed research; E.V.[2] purified samples for mass spectrometry; L.B., M.P., P.H., D.B., F.S., A.D., and H.R.M. performed mass spectrometric analysis and data interpretation; E.V.[2], A.A., K.F., and S.T. purified samples for NMR; E.V.[3], A.A., K.F., S.T., and S.M.L. performed NMR analysis and data interpretation. L.B. and S.M.L. prepared figures; L.B., M.P., E.V.[3], S.T., S.M.L., B.W.W., A. D., and H.R.M. wrote the manuscript; all authors edited the paper.

……………………………………………………………………………………………………

References:

1. He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., Connor, T. R., Harris, S. R., Fairley, D., Bamford, K. B., D'Arc, S., Brazier, J., Brown, D., Coia, J. E., Douce, G., Gerding, D., Kim, H. J., Koh, T. H., Kato, H., Senoh, M., Louie, T., Michell, S., Butt, E., Peacock, S. J., Brown, N. M., Riley, T., Songer, G., Wilcox, M., Pirmohamed, M., Kuijper, E., Hawkey, P., Wren, B. W., Dougan, G., Parkhill, J., and Lawley, T. D. (2013) Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nature Genetics* **45**, 109-113

2. Cairns, M. D., Stabler, R. A., Shetty, N., and Wren, B. W. (2012) The continually evolving *Clostridium difficile* species. *Future Microbiology* **7**, 945-957

3. Kirby, J. M., Ahern, H., Roberts, A. K., Kumar, V., Freeman, Z., Acharya, K. R., and Shone, C. C. (2009) Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*. *J. Biol. Chem.* **284**, 34666-34673

4. Eidhin, D. N., Ryan, A. W., Doyle, R. M., Walsh, J. B., and Kelleher, D. (2006) Sequence and phylogenetic analysis of the gene for surface layer protein, *slpA*, from 14 PCR ribotypes of *Clostridium difficile*. *J. Med. Microbiol* **55**, 69-83

5. Hennequin, C., Janoir, C., Barc, M. C., Collignon, A., and Karjalainen, T. (2003) Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology* **149**, 2779-2787

6. Reynolds, C. B., Emerson, J. E., de la Riva, L., Fagan, R. P., and Fairweather, N. F. (2011) The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS Pathog.* **7**, e1002024

7. Barketi-Klai, A., Monot, M., Hoys, S., Lambert-Bordes, S., Kuehne, S. A., Minton, N., Collignon, A., Dupuy, B., and Kansau, I. (2014) The Flagellin FliC of *Clostridium difficile* Is Responsible for Pleiotropic Gene Regulation during In Vivo Infection. *PLoS ONE* **9**, e96876

8. Twine, S. M., Reid, C. W., Aubry, A., McMullin, D. R., Fulton, K. M., Austin, J., and Logan, S. M. (2009) Motility and Flagellar Glycosylation in *Clostridium difficile*. *J. Bacteriol.* **191**, 7050-7062

9. Faulds-Pain, A., Twine, S. M., Vinogradov, E., Strong, P. C. R., Dell, A., Buckley, A. M., Douce, G. R., Valiente, E., Logan, S. M., and Wren, B. W. (2014) The post-translational modification of the *Clostridium difficile* flagellin affects motility, cell surface properties and virulence. *Mol Microbiol.* **94**, 272-289

10. Merino, S., and Tomás, J. M. (2014) Gram-Negative Flagella Glycosylation. *Int. J. Mol. Sci.* **15**, 2840-2857

11. Schirm, M., Kalmokoff, M., Aubry, A., Thibault, P., Sandoz, M., and Logan, S. M. (2004) Flagellin from *Listeria monocytogenes* Is Glycosylated with β-O-Linked N-Acetylglucosamine. *J. Bacteriol.* **186**, 6721-6727

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

13. Janesch, B., Schirmeister, F., Maresch, D., Altmann, F., Messner, P., Kolarich, D., and Schäffer, C. (2016) Flagellin glycosylation in *Paenibacillus alvei* CCM 2051T. *Glycobiology* **26**, 74-87

14. Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeno-Tarraga, A. M., Wang, H., Holden, M. T., Wright, A., Churcher, C., Quail, M. A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B.,

Dougan, G., Barrell, B., and Parkhill, J. (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics* **38**, 779-786

15. Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J., and Bateman, R. H. (1996) High Sensitivity Collisionally-activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry* **10**, 889-896

16. Morris, H. R., Paxton, T., Panico, M., McDowell, R., and Dell, A. (1997) A Novel Geometry Mass Spectrometer, the Q-TOF, for Low-Femtomole/Attomole-Range Biopolymer Sequencing. *J. Protein Chem.* **16**, 469-479

17. Hunt, E., and Morris, H. R. (1973) Collagen cross-links. A mass-spectrometric and (1)H- and (13)C-nuclear-magnetic-resonance study. *Biochem. J.* **135**, 833-843

18. Billker, O., Lindo, V., Panico, M., Etienne, A. E., Paxton, T., Dell, A., Rogers, M., Sinden, R. E., and Morris, H. R. (1998) Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature* **392**, 289-292

19. Szymanski, C. M., and Wren, B. W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Micro.* **3**, 225-237

20. Asakura, H., Churin, Y., Bauer, B., Boettcher, J. P., Bartfeld, S., Hashii, N., Kawasaki, N., Mollenkopf, H. J., Jungblut, P. R., Brinkmann, V., and Meyer, T. F. (2010) *Helicobacter pylori* HP0518 affects flagellin glycosylation to alter bacterial motility. *Mol. Microbiol.* **78**, 1130-1144

21. Iwashkiw, J. A., Seper, A., Weber, B. S., Scott, N. E., Vinogradov, E., Stratilo, C., Reiz, B., Cordwell, S. J., Whittal, R., Schild, S., and Feldman, M. F. (2012) Identification of a General O-linked Protein Glycosylation System in *Acinetobacter baumannii* and Its Role in Virulence and Biofilm Formation. *PLoS Pathog.* **8**, e1002758

22. Guerry, P., Ewing, C. P., Schirm, M., Lorenzo, M., Kelly, J., Pattarini, D., Majam, G., Thibault, P., and Logan, S. (2006) Changes in flagellin glycosylation affect Campylobacter autoagglutination and virulence. *Mol. Microbiol.* **60**, 299-311

23. Alemka, A., Nothaft, H., Zheng, J., and Szymanski, C. M. (2013) N-Glycosylation of Campylobacter jejuni Surface Proteins Promotes Bacterial Fitness. *Infect. Immun.* **81**, 1674-1682

24. Szymanski, C. M., Burr, D. H., and Guerry, P. (2002) Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect. Immun.* **70**, 2242-2244

25. Larsen, J. C., Szymanski, C., and Guerry, P. (2004) N-Linked Protein Glycosylation Is Required for Full Competence in *Campylobacter jejuni* 81-176. *J. Bacteriol.* **186**, 6508-6514

26. Valiente, E., Faulds-Pain, A., Hitchen, P., Bouché, L., Songane, M., Stabler, R., Dawson, L., Donahue, E., Panico, M., Morris, H.R., Bajaj-Elliott, M., Logan, S., Dell, A. and Wren, B. (2016) Role of glycosyltransferases modifying type B flagellin of emerging hypervirulent *Clostridium difficile* and their impact on motility and adhesion. *Companion paper; under review J. Biol. Chem.*

27. Logan, S. M. (2006) Flagellar glycosylation – a new component of the motility repertoire? *Microbiology* **152**, 1249-1262

28. Arnold, F., Bédouet, L., Batina, P., Robreau, G., Talbot, F., Lécher, P., and Malcoste, R. (1998) Biochemical and Immunological Analyses of the Flagellin of *Clostridium tyrobutyricum* ATCC 25755. *Microbiology and Immunology* **42**, 23-31

29. Bedouet, L., Arnold, F., Robreau, G., Batina, P., Talbot, F., and Binet, A. (1998) Evidence for an heterogeneous glycosylation of the *Clostridium tyrobutyricum* ATCC 25755 flagellin. *Microbios* **94**, 183-192

30. Lyristis, M., Boynton, Z.L., Petersen, D., Kan, Z., Bennett, G.N., and Rudolph, F.B. (2000) Cloning, sequencing, and characterization of the gene encoding flagellin, flaC, and the post-translational modification of flagellin, FlaC, from *Clostridium acetobutylicum* ATCC824. *Anaerobe* 69-79

31. Schirm, M., Schoenhofen, I. C., Logan, S. M., Waldron, K. C., and Thibault, P. (2005) Identification of Unusual Bacterial Glycosylation by Tandem Mass Spectrometry Analyses of Intact Proteins. *Anal. Chem.* **77**, 7774-7782

32. Miao, J., Fa, Y., Gu, B., Zhu, W., and Zou, S. (2012) Taurine attenuates lipopolysaccharide-induced disfunction in mouse mammary epithelial cells. *Cytokine* **59**, 35-40

33. Dorrell, N., Mangan, J. A., Laing, K. G., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B. G., Parkhill, J., Stoker, N. G., Karlyshev, A. V., Butcher, P. D., and Wren, B. W. (2001) Whole Genome Comparison of *Campylobacter jejuni* Human Isolates Using a Low-Cost Microarray Reveals Extensive Genetic Diversity. *Genome Res.* **11**, 1706-1715

34. Pfoestl, A., Hofinger, A., Kosma, P., and Messner, P. (2003) Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-α-D-galactose in *Aneurinibacillus thermoaerophilus L420-91T. J. Biol. Chem.* **278**, 26410-26417

35. Naundorf, A., and Klaffke, W. (1996) Substrate specificity of native dTDP-D-glucose-4,6-dehydratase: chemo-enzymatic syntheses of artificial and naturally occurring deoxy sugars. *Carbohyd. Res.* **285**, 141-150

Clostridium difficile type B flagellin is glycosylated with sulphonated peptidylamido-glycans

Table 1: Atomic compositions deduced for key signals in the high resolution MS/MS data.

This table shows the deduced atomic compositions for the experimental (measured) masses observed for key signals in the high resolution MS/MS data (Fig 1) together with the theoretical masses of those compositions. Note that certain fragments, such as peptide y" ions and the peptide quasimolecular ion (M+H$^+$) serve as useful internal standards, confirming the mass accuracies across the data set. Key discoveries from these data included the finding of sulphur in the m/z 152 and higher mass fragments, thus confirming a novel structural unit and allowing the interpretation of a clear fragmentation pathway between the m/z 396 and 152 ions. For a full interpretation see the text and Scheme 1.

Observed Mass (m/z)	Atomic Composition Assigned	Theoretical Mass (m/z
111.0441	C6H7O2	111.0446
152.0383	C4H10NO3S	152.0381
175.1192	y"1	175.1195
204.0873	C8H14NO5	204.0872
223.0753	C7H15N2O4S	223.0752
251.0700	C8H15N2O5S	251.0701
268.0965	C8H18N3O5 S	268.0967
288.2034	y"2	288.2035
378.1333	C14H24N3O7S	378.1335
396.1432	C14H26N3O8S	396.1441
749.3781	y"7	749.3793
952.4579	Y"7 + HexNAc	952.4587
1090.5752	Peptide (M + H+)	1090.5745
1293.6548	Peptide + HexNAc	1293.6538

Table 2. NMR data for compound 1. NAc: 175.8; 2.11/23.7 ppm. Me: 3.44/57.2; 3.45/57.7 ppm

Fr. 21	H/C-1	H/C-2	H/C-3	H/C-4	H/C-5	H/C-6
RhaMe	5.08	3.97	3.56	3.50	4.02	1.25
	100.9	76.8	80.4	72.2	70.2	17.6
RhaMe	4.89	4.27	3.45	3.46	3.75	1.29
	103.5	67.0	80.5	72.2	70.1	18.2
GlcNAc	4.67	3.87	3.60	3.54	3.49	3.79; 3.94
	100.9	56.3	82.9	69.5	77.1	61.7
Ser*		4.36	4.17; 4.17			
	168.2	54.2	68.0			

14

Table 3. NMR data for compounds 2 and 3. Atom numbering in HSO$_3$CH$_2$CH$_2$NH-Ala (X) from Ala side.

NAc: 175.8; 2.11/23.7 ppm. Me: 3.44/57.2; 3.45/57.7 ppm

Fr. 21	H/C-1	H/C-2	H/C-3	H/C-4	H/C-5	H/C-6
RhaMe	5.08	3.97	3.56	3.51	4.01	1.25
	101.0	77.1	80.4	72.2	70.3	17.6
Rha	4.89	4.21	3.80	3.58	3.80	1.32
	103.2	68.2	76.9	71.4	70.2	18.3
GlcNAc	4.66	3.87	3.60	3.54	3.48	3.77; 3.94
	100.9	56.3	83.0	69.5	77.1	61.8
Fuc3NR	5.09	3.88	4.31	3.78	4.42	
	96.6	67.1	52.4	71.6	68.0	
Ser*		4.36	4.17; 4.17			
		54.2	68.0			
Ala	171.5	4.20	1.61			
		57.5	16.5			
Gly		4.06				
		43.7				
X	3.47; 3.51	3.32; 3.32				
	42.9	47.7				
RhaMe	4.89	4.27	3.45	3.46	3.74	1.29
	103.2	67.0	80.5	72.2	70.1	18.2
RhaMe	5.09	3.99	3.56	3.51	4.01	1.25
	100.9	76.5	80.4	72.2	70.3	17.6
GlcNAc	4.62	3.79	3.60	3.50	3.45	3.75; 3.92
	100.2	56.8	83.0	69.7	77.0	62.0
Thr*		4.53	4.32			
		59.3	75.9			

Clostridium difficile type B flagellin is glycosylated with sulphonated peptidylamido-glycans

Table 4. Strains used in this study

Strains	Characteristics	Source
Clostridium difficile		
R20291	PCR ribotype 027, isolated from an outbreak in 2004-2005	Stabler et al., 2009
CD196	ribotype 027,isolated from a pseudomembranous colitis case	France, 1985
BI-16	PCR ribotype 027, isolated from an outbreak in 2004	Augusta, USA
CD305	PCR ribotype 023, isolated from an outbreak in 2011	Barts Hospital, UK
CD1426	PCR ribotype 023, isolated from an outbreak in 2010	Queens Hospital Remford, UK
CD1714	PCR ribotype 023, isolated from an outbreak in 2011	Whipp's Cross Hospital, UK
106-01	PCR ribotype 106, isolated from an outbreak in 2008	Fatal case, Glasgow, UK
001-01	PCR ribotype 001, isolated from an outbreak in 2008	Paisley, UK
001-07	PCR ribotype 001, isolated from an outbreak in 2008	Edinburgh, UK

16

LEGENDS

Figure 1: Positive ion on-line nano-LC MS/MS high resolution CAD mass spectrum of m/z 998^{2+} (main picture) with equivalent low resolution MS/MS spectrum of m/z 991^{2+} (inset). Note the low-mass signals (below m/z 400) many of which do not correlate with either known peptide or carbohydrate fragments, and which therefore indicated the discovery of novel structural features in these glycopeptides. For full interpretation see the text and Scheme 1.

Scheme 1: Summary of the interpretation and mechanistic logic used to assess the mass spectrometric fragmentation data determined in this study (see Fig 1) showing the probable structural assignments. Conclusions were aided by the derivation of atomic compositions of key fragments shown in Table 1.

Figure 2: Positive ion nanospray CAD MS/MS spectrum of m/z 152 produced via cone-voltage induced in-source fragmentation of m/z 998^{2+}. Signals at m/z 70 and 108 correspond to losses of 82 and 44 mass units respectively, and were correlated with the low-mass high resolution data from Fig. 1 to deduce atomic compositions for these ions of C_4H_8N and $C_2H_6NO_2S$. This suggests two overlapping fragments derived from the m/z 152 ion ($C_4H_{10}NO_3S$ in Table 1), an alkylamine and a possible sulphonic acid. For a full interpretation see the text and Scheme 1.

Figure 3: Negative ion nanospray CAD MS/MS spectrum of m/z 996^{2-} (m/z 100-1200 mass range). Negative ion fragmentation data complement and expand that from positive ion experiments, as here for m/z 394 (equivalent to m/z 396 in Fig.1), m/z 266 (m/z 268), m/z 173 (m/z 175) and m/z 150 (m/z 152). Interestingly, a moderately intense low-mass fragment was observed at m/z 124 (lowest mass significant fragment) and this was therefore chosen for further CAD MS/MS analysis – see Fig 4. The highest mass signal seen at m/z 921 corresponds to the elimination of the complete sulphonated peptidylamido-glycan from the peptide backbone to leave a dehydropeptide (not observed). For a full interpretation see the text.

Figure 4: The negative ion nanospray CAD MS/MS spectrum of m/z 124 produced via cone voltage induced in-source fragmentation of m/z 996^{2-}. The base peak (main signal) observed is present at m/z 79.96, interpreted as SO_3^- together with minor, less informative signals. This spectrum provided the first evidence of the presence of a sulphonic acid group in the new structure. A sample of synthetic Taurine was then analysed, showing an equivalent negative ion spectrum with base peak m/z 79.96.

Figure 5: *C. difficile* R20291 flagellar glycan compounds. Structures of isolated compound 1, 2 and 3 are presented. Letters A, B, C, C', D, E and F below the structures are labelled according to the assignment of resonances presented in Fig 6 and Tables 2 and 3. 381 indicates the fragment identified by MS analysis in Fig 1 (inset m/z 382).

Figure 6: Part of the ^1H-^{13}C HSQC spectrum of the mixture of compounds 2 and 3. Signal marked # probably belongs to a structure 2 with additional Me group, not yet localized. Amino acids marked with superscript numbers are from peptide part, sequence not determined. Thr* and Ser* are glycosylated.
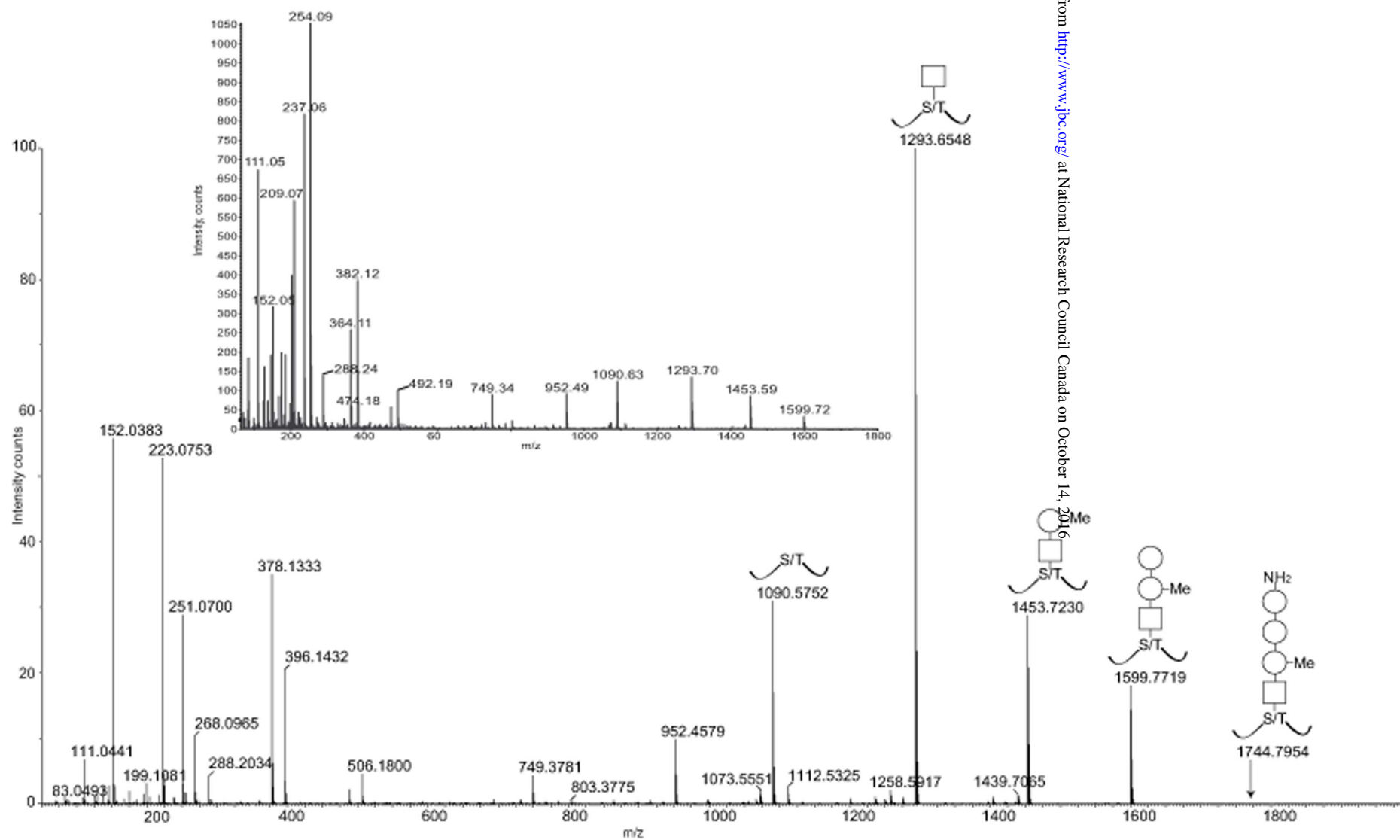
Clostridium difficile type B flagellin is glycosylated with sulphonated peptidylamido-glycans
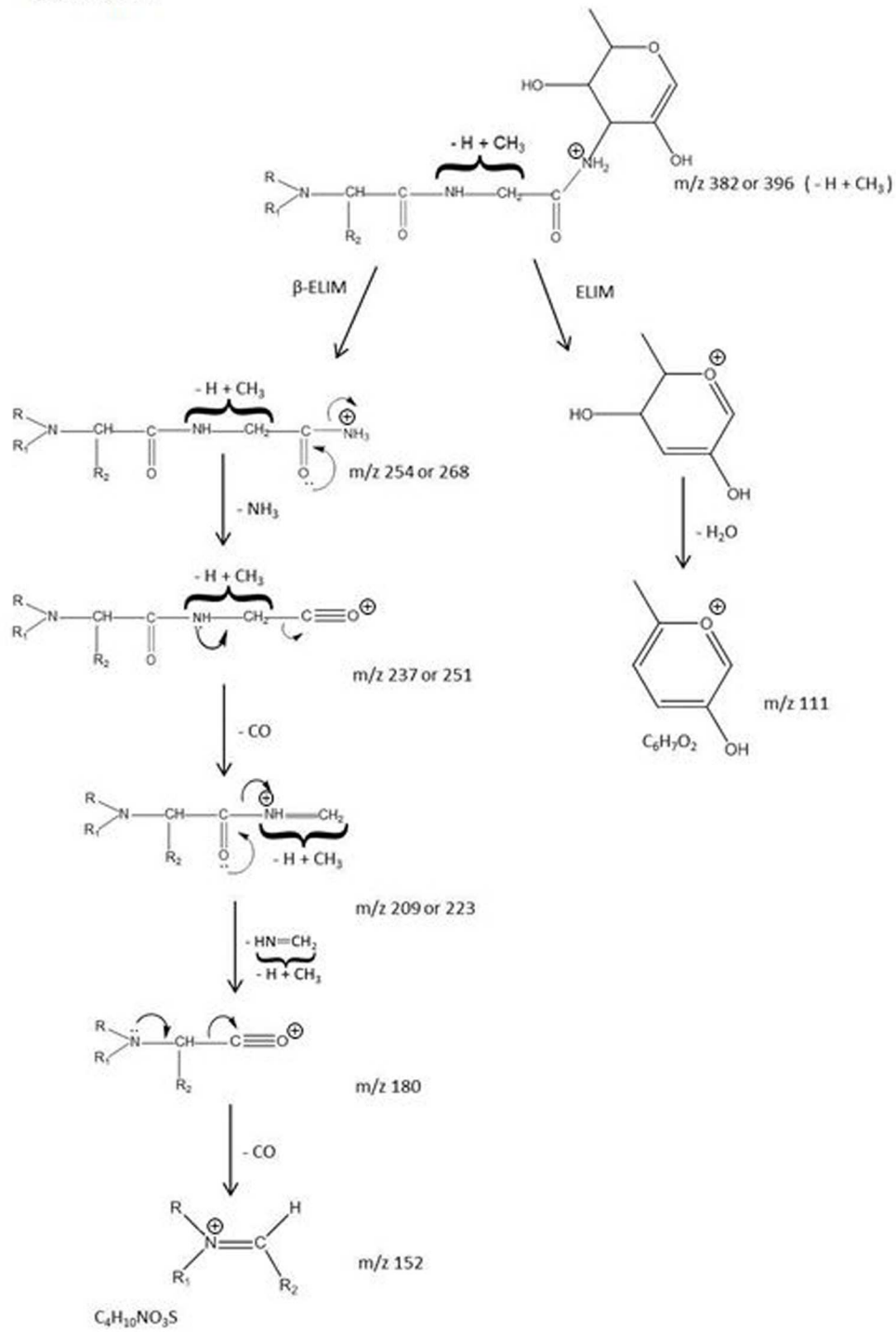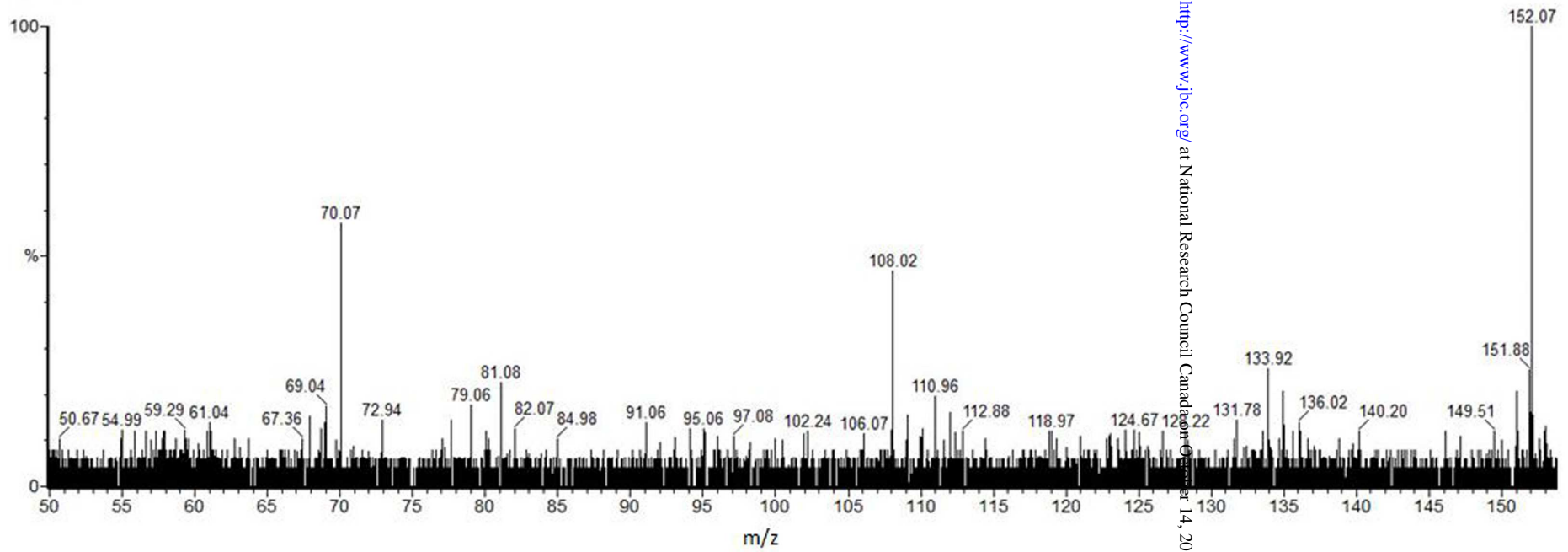

Figure 1

Scheme 1

m/z 382 or 396 (- H + CH₃)

β-ELIM

ELIM

m/z 254 or 268

- NH₃

- H₂O

m/z 237 or 251

- CO

m/z 111

$C_6H_7O_2$

m/z 209 or 223

- HN=CH₂
- H + CH₃

m/z 180

- CO

m/z 152

$C_4H_{10}NO_3S$

Figure 2

Figure 3

Figure 4

Figure 5

α-Rha3OMe-2-α-Rha3OMe-3-β-GlcNAc-Ser* **1**
   B   A   C

Taurine   Ala   Gly

O—3-α-Rha-2-α-Rha3OMe-3-β-GlcNAc-3-Ser* **2**
       E   A   C

F (Fuc3NR)

α-Rha3OMe-2-α-Rha3OMe-3-β-GlcNAc-Thr* **3**
   B   D   C'

23

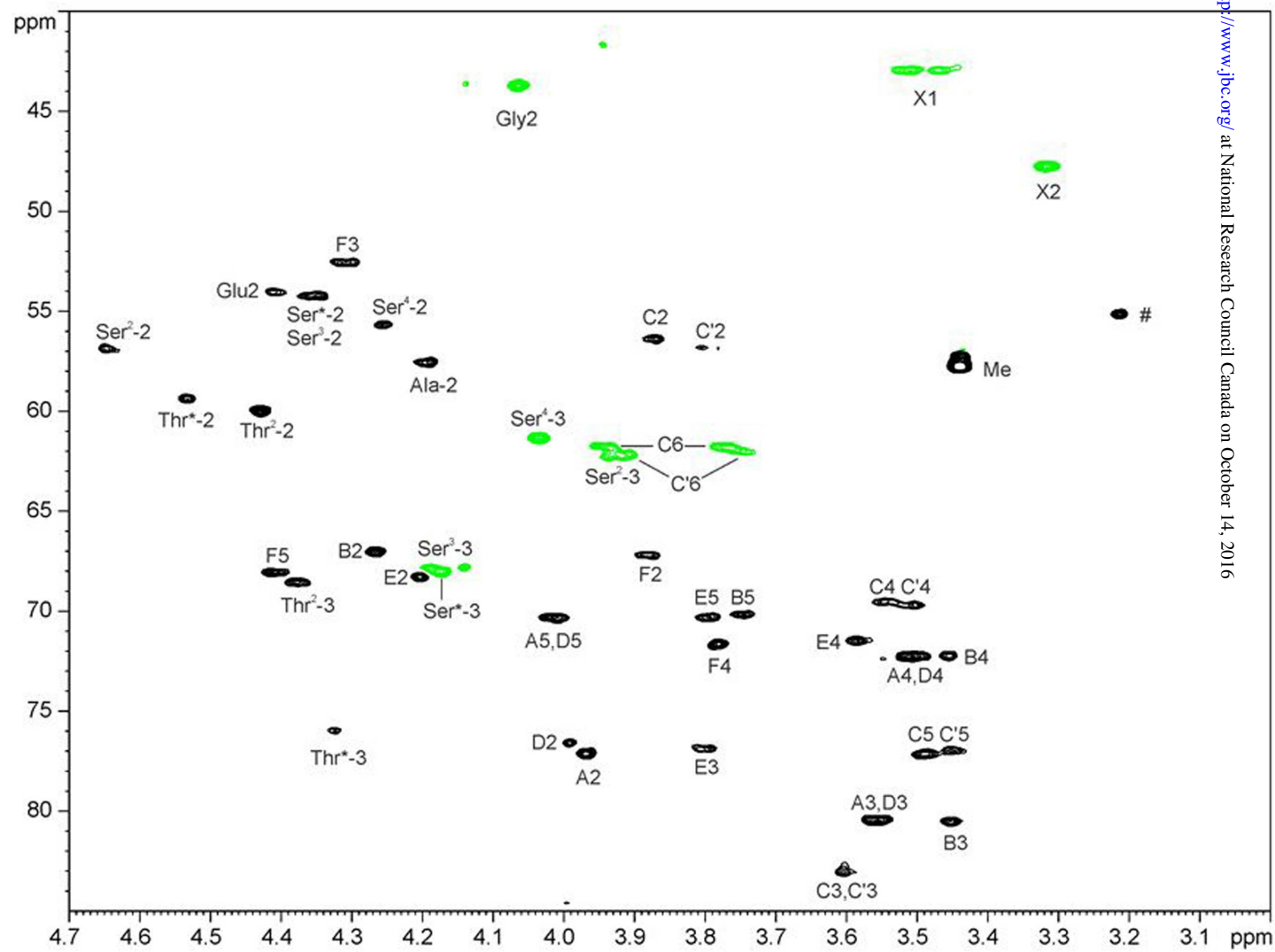*Clostridium difficile* type B flagellin is glycosylated with sulphonated peptidylamido-glycans
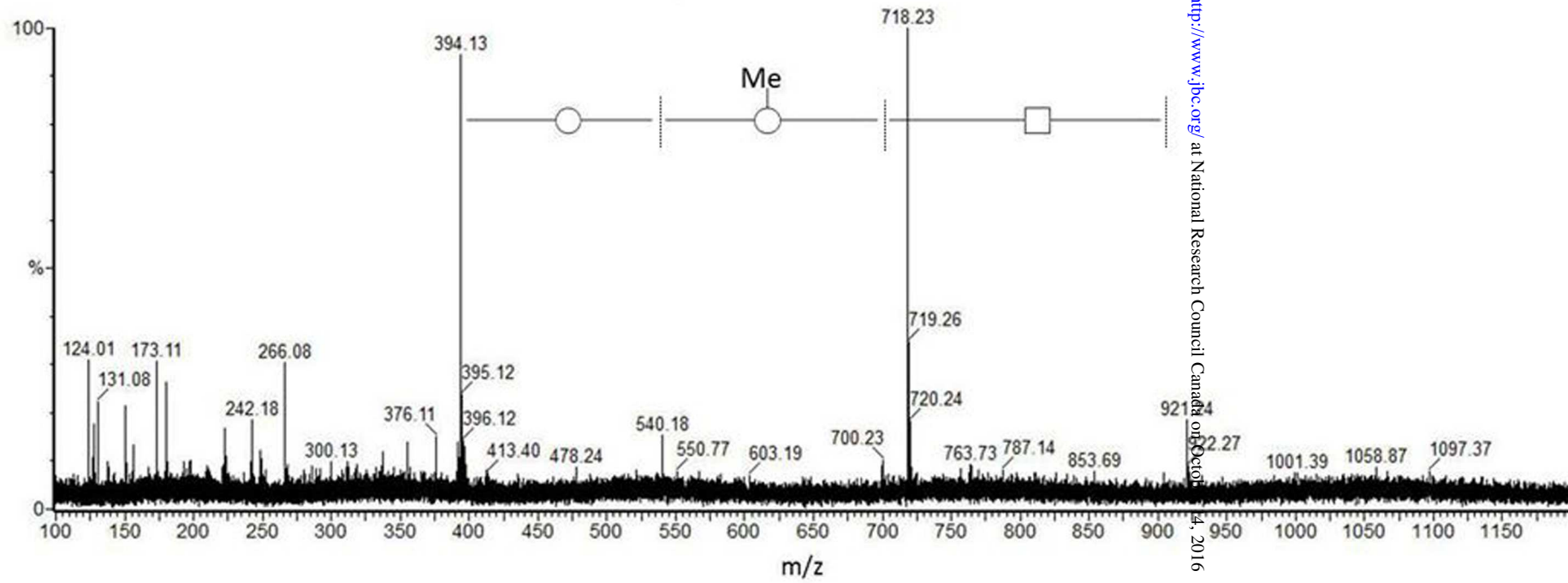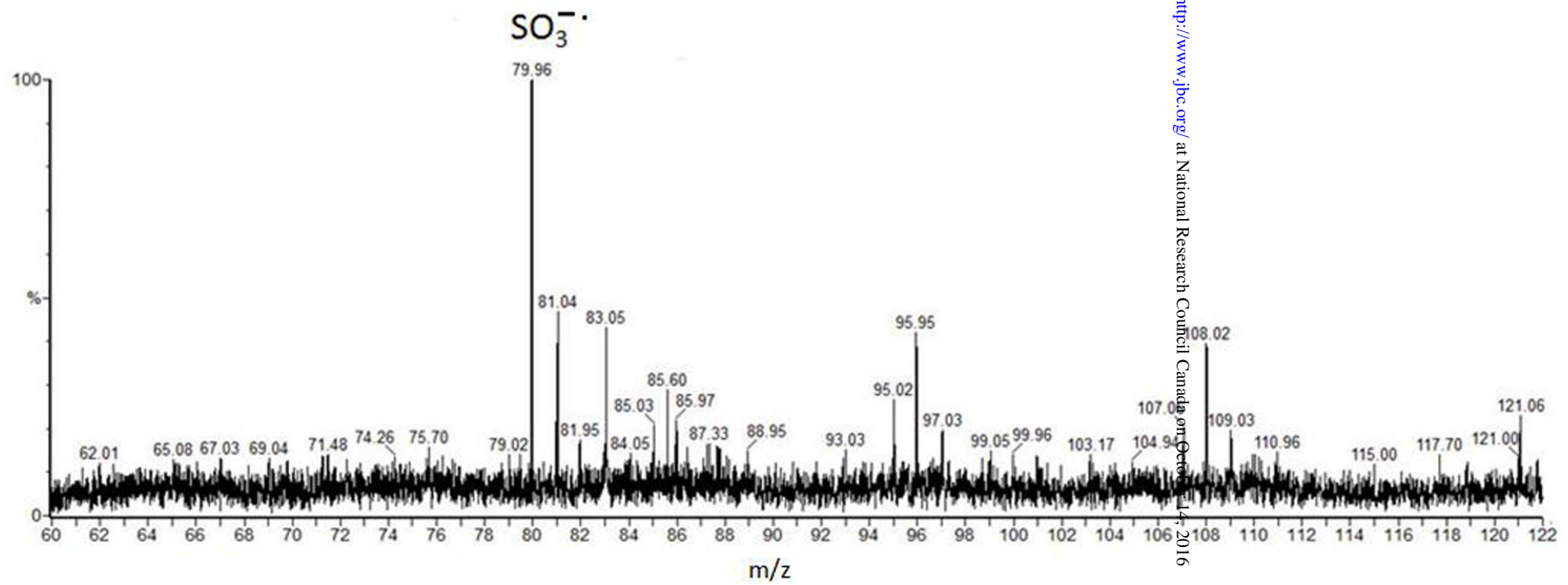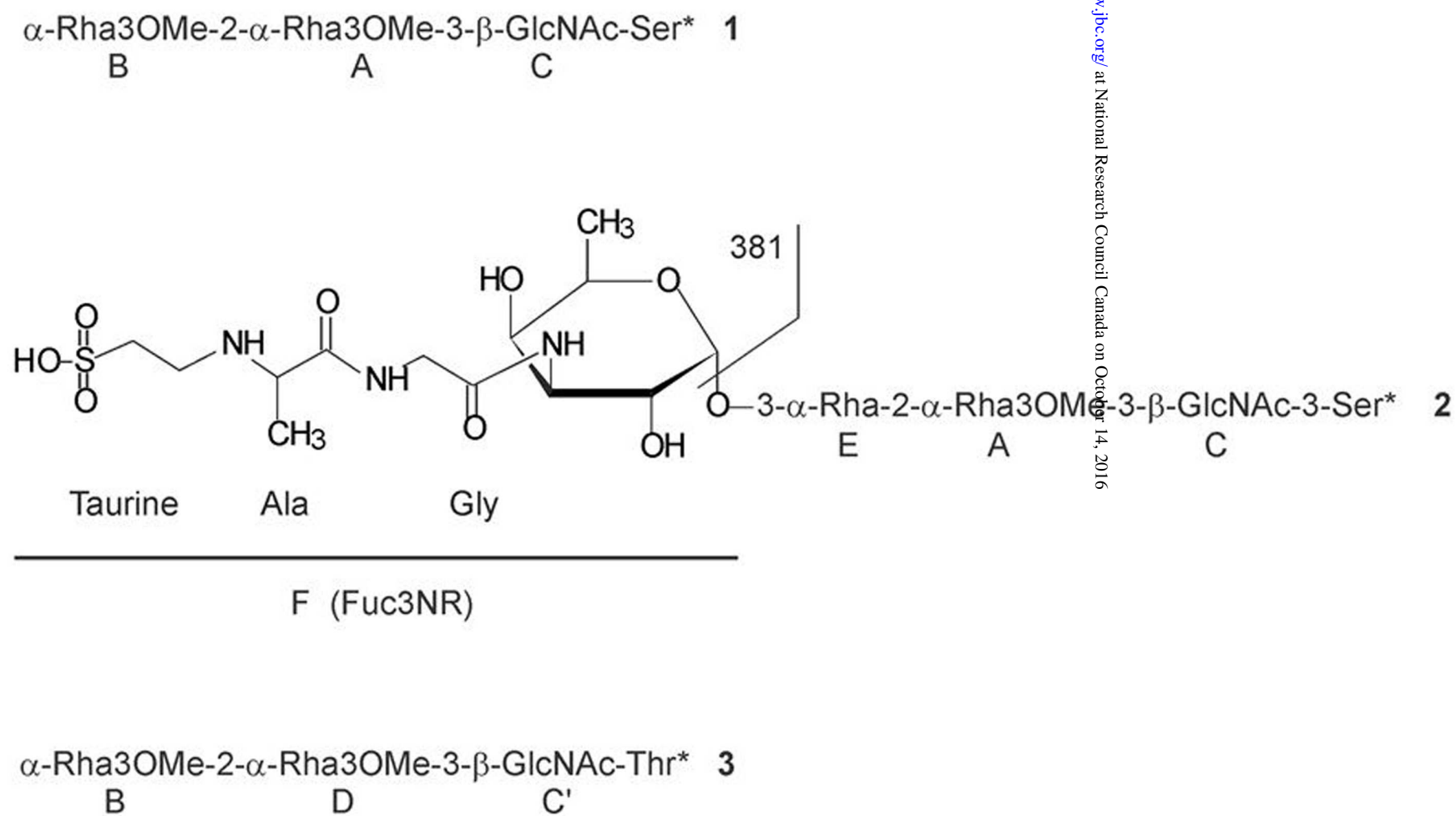
Figure 6

The Type B Flagellin of Hypervirulent Clostridium difficile is Modified with Novel Sulphonated Peptidylamido-Glycans

Laura Bouché, Maria Panico, Paul Hitchen, Daniel Binet, Federico Sastre, Alexandra Faulds-Pain, Esmeralda Valiente, Evgeny Vinogradov, Annie Aubry, Kelly Fulton, Susan Twine, Susan M Logan, Brendan W Wren, Anne Dell and Howard R Morris

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2016/10/07/jbc.M116.749481.full.html#ref-list-1