Identification of the Carbohydrate Moieties and Glycosylation Motifs in Campylobacter jejuni Flagellin*

Pierre Thibault‡, Susan M. Logan‡§, John F. Kelly‡, Jean-Robert Brisson‡, Cheryl P. Ewing‡, Trevor J. Trust†, and Patricia Guerry¶

From the ‡Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada, †Division of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and ¶Naval Medical Research Center, Silver Spring, Maryland 20910

Flagellins from three strains of Campylobacter jejuni and one strain of Campylobacter coli were shown to be extensively modified by glycosyl residues, imparting an approximate 6000-Da shift from the molecular mass of the protein predicted from the DNA sequence. Tryptic peptides from C. jejuni 81–176 flagellin were subjected to capillary liquid chromatography-electrospray mass spectrometry with a high/low orifice stepping to identify peptide segments of aberrant masses together with their corresponding glycosyl appendages. These modified peptides were further characterized by tandem mass spectrometry and preparative high performance liquid chromatography followed by nano-NMR spectroscopy to identify the nature and precise site of glycosylation. These analyses have shown that there are 19 modified Ser/Thr residues in C. jejuni 81–176 flagellin. The predominant modification found on C. jejuni flagellin was O-linked 5,7-diacetamido-3,5,7,9-tetraoxo-L-glycero-1-manno-nonulosonic acid (pseudaminic acid, Pse5Ac7Ac) with additional heterogeneity conferred by substitution of the acetamido groups with acetamidino and hydroxypropionyl groups. In C. jejuni 81–176, the gene Cj1316c, encoding a protein of unknown function, was shown to be involved in the biosynthesis and/or the addition of the acetamido group on Pse5Ac7Ac. Glycosylation is not random, since 19 of the total 107 Ser/Thr residues are modified, and all but one of these are restricted to the central, surface-exposed domain of flagellin when folded in the filament. The mechanism of attachment appears unrelated to a consensus peptide sequence but is rather based on surface accessibility of Ser/Thr residues in the folded protein.

Campylobacter spp. are among the most frequent causative agents of bacterial diarrhea worldwide and the leading cause of food-borne illness in North America (1, 2). Motility is an essential virulence determinant required for colonization of the gastrointestinal tract and invasion of intestinal epithelial cells in vitro (3). Moreover, Campylobacter jejuni flagellin is the immunodominant protein recognized during infection and has been suggested to be an immunoprotective antigen (4–7). The flagellar filaments of Campylobacter spp. are complex, composed primarily of the FlaA flagellin but with trace amounts of a highly homologous flagellin, FlaB (7). Flagellins from numerous strains of C. jejuni and the related organism, Campylobacter coli, have been shown to be glycosylated (8), and the modifications have been shown to occur on ≥13 serine residues on flagellin from strain C. coli (9). In addition, the glycosyl modifications are surface-exposed in the flagellar filament and appear to be highly immunogenic (10). Several genes involved in glycosylation of Campylobacter flagellin have been described (11–13). Two of these genes encode homologs of prokaryotic enzymes, sialic acid (Neu5Ac) synthase and CMP-Neu5Ac synthetase, which are involved in synthesis of Neu5Ac (11, 12). These observations coupled with reports that a Neu5Ac-specific lectin can bind to Campylobacter flagellins, have led to the hypothesis that the glycosyl posttranslational modifications on flagellin include Neu5Ac moieties. Although protein glycosylation was previously considered to be restricted to eukaryotes and archaea, there are increasing examples of prokaryotic glycoproteins (14–16) including pilins from Neisseria spp. (17) and Pseudomonas aeruginosa (18) and flagellins from not only Campylobacter spp., but also Caulobacter crescentus (19) and P. aeruginosa (20). We have undertaken a comprehensive structural analysis of C. jejuni and C. coli flagellin by mass spectrometry and NMR spectrometry to identify the precise nature of these modifications and the potential structural determinants underlying the selective glycosylation of Campylobacter flagellin. Furthermore, we have determined that a gene encoding a protein of unknown function in the C. jejuni genome sequence is involved in biosynthesis of one of the modifications.

EXPERIMENTAL PROCEDURES

Bacterial Strains—C. jejuni strains 81–176, NCTC 11168, OH4384 and C. coli strain VC167 T2 have been described previously (9, 21–23). Purification of Flagellin and Preparation of Tryptic Peptides—Campylobacter strains were grown in Mueller Hinton broth overnight at 37 °C under microaerobic conditions in batches of 1.6 liters. Flagellin was purified by the method of Power et al. (10). Approximately 2 mg of flagellin was digested overnight at 37 °C with trypsin (Promega, Madison WI) and subsequently evaporated using a Speedvac preconcentrator. Preparative HPLC separations were conducted on a 25 × 1-cm C18 column.

* This work was supported by NIAID, National Institutes of Health Grant AI43559 (to P. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF345999 and AF346000.

†§ To whom correspondence should be addressed: Institute of Biological Sciences, 100 Sussex Dr., Ottawa, Ontario K1A 0R6, Canada. Tel.: 613-990-0839; Fax: 613-952-9092; E-mail: susan.logan@nrc.ca.

¶ The abbreviations used are: HPLC, high performance liquid chromatography; ELC-ESMS, capillary liquid chromatography electrospray mass spectrometry; MS-MS, tandem mass spectrometry; COSY, correlated spectroscopy; NOE, nuclear Overhauser effect; Pse5Ac7Ac, pseudaminic acid, 5,7-diacetamido-3,5,7,9-tetraoxo-L-glycero-1-manno-nonulosonic acid; Pse5Ac7Am7Ac, 5-acetamidino-7-acetamido-L-glycero-1-manno-nonulosonic acid; Pse5Am7Ac, 5-acetamido-7-acetamido-L-glycero-1-manno-nonulosonic acid; Pse5Ac7Am7Ac, 5,7-diacetamido-8-O-acetyl-Pse; OR, orifice; CID, collision-induced dissociation; IEF, isoelectric focusing; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; LPS, lipopolysaccharide.
Vydac 218TP510 column (Hesperia, CA) using a gradient of 5–90% aqueous acetonitrile (0.1% trifluoroacetic acid) for 30 min. The UV detector was set to 214 nm. Replicate injections of 200 µg each were made on the column, and fractions were collected every 1 min over the course of the gradient elution. For β-elimination experiments, purified tryptic glycopeptides (previously identified by mass spectrometry) were incubated for periods of 6–16 h in a 25% aqueous solution of ammonium hydroxide. Samples were then evaporated to dryness on a Speedvac preconcentrator and reconstituted in water, and remaining salts were removed by passing the solution on a Millipore Ziptip C18 (Bedford, MA) and eluting the peptide using 50% aqueous methanol solution (0.2% formic acid).

**Mass Spectrometry**—All mass spectra were obtained on a PerkinElmer/Sciex Q-Star mass spectrometer (Concord, ON, Canada) using liquid chromatography Tune and Biomultiview programs for data acquisition and processing, respectively. An HP 1100 liquid chromatograph was coupled to the Q-Star for LC-MS experiments. Capillary LC separations were conducted with a 15-cm × 0.32-mm Pepmap C18 column (LC Packings, San Francisco, CA) using a linear gradient of 5–95% aqueous acetonitrile (0.2% formic acid) for 30 min. A pre-column flow splitter was mounted before a Rhodyne 8125 injector to provide a column flow rate of 3.5 µL/min. Injections of typically 1 µg of flagellin tryptic digests were made on the capillary column.

For conventional mass spectra of intact flagellin, ~1 µg of purified protein in 0.2% formic acid was flow-injected into a stream of 50% aqueous acetonitrile (0.2% formic acid). The mass spectrometer was set to record the range m/z 1000–2500. During the LC-ESMS experiments, a stepped orifice voltages ramp was set over two distinct scanning periods (period 1: 1 s, stepped orifice (OR) 100 V m/z 150–400; period 2: 2 s, OR 30 V, m/z 400–18000) for the acquisition of conventional mass spectra. Tandem mass spectrometry experiments were conducted on the Q-Star using a nanoelectrospray interface. HPLC fractions of tryptic glycopeptides before and after β-elimination were loaded in the open end of a nanoelectrospray type A emitter (Micromass, Manchester, UK). Collision-induced dissociation (CID) of selected precursor ions identified in a preliminary survey scan was achieved using nitrogen as the collision gas at collision energies of typically 50–100 eV (laboratory frame of reference). Fragment ions formed in the rf-only quadrupole were recorded by the time of flight mass analyzer.

**NMR Spectroscopy**—NMR experiments were performed on a Varian INOVA 600 NMR spectrometer with VNMR 6.1B software using a gradient inverse broadband nano-NMR probe at a spin rate of 2600 Hz (23, 24). HPLC-purified samples (200–500 µg) were evaporated on a Speedvac concentrator and redisolved in 40 µL of D2O with no adjustment of pH. NMR experiments were done at 25°C with the HOD resonance set at 4.77 ppm. All experiments were performed as described before (25). For the analysis of the NOE data, coordinates for 5-acetamido-7-acetamido-3,5,7,9-tetradeoxy-t-glycero-α-D-manno-nonulosonic acid were generated using MM3(92) (QCPE) and Insight II (Molecular Simulations Inc.). The dihedral angles about the C6–C7–C8–C9 bonds were based on the major conformers found in solution (26, 27).

**Cloning and Genetic Analysis of Flagellin Modification Genes**—The 81–176 homolog of Cj1317 and adjacent DNA was cloned from a λ-ZAP Express library using as probe a polymerase chain reaction product generated from the Cj1317 sequence (22). DNA sequencing was done with dye terminator chemistry on an Applied Biosystems Model 373A sequencer. Mutants were constructed using an in vitro Tn5-based transposition system EZTn pMOD (Epitrend, Madison, WI) containing a Campylobacter chloramphenicol (Cm’)-resistance gene from pRY109 as previously described (28). The transposon was polymerase chain reaction-amplified with primers specified by Epitrend and used in an in vitro transposition reaction with the target plasmid. The reaction was transformed into Escherichia coli DH5α, and plasmid DNAs from individual transformants were sequenced using primers within the Cm’ cassette to determine the insertion point and orientation with respect to the target gene. Insertions into Cj1314c, Cj1315c, Cj1316c, and Cj1317 were electroporated into 81–176 (29) with selection on Mueller Hinton agar supplemented with 15 µg/ml chloramphenicol under microaerobic conditions. DNAs from individual Cm’ colonies were subjected to polymerase chain reaction using primers bracketing the insertion site of each insertion to confirm that the mutated allele had integrated into the 81–176 chromosome by a double crossover.

To perform complementation in trans with the Cj1316c mutant, DNA encoding the wild type alleles from plasmid pSG1854 was subcloned from the excision plasmid of λ-ZAP into the kanamycin resistant (Km’)-shuttle plasmid, pRY107 (30), to generate pRY107/1854. The construction was transformed into E. coli DH5α containing plasmid RK212.2 (31). Plasmid pRY107/1854 was conjugally mobilized by RK212.2 from E. coli into the Cj1316c mutant of 81–176 with selection on Mueller Hinton agar supplemented with 10 µg/ml trimethoprim, 15 µg/ml chloramphenicol, and 25 µg/ml kanamycin. As a control pRY107 was also transferred into the Cj1316c mutant.

**Motility Testing**—Motility of mutants was compared with that of wild type on semi-solid (0.4%) Mueller Hinton agar plates as previously described (4).

**Isoelectric Focusing (IEF)**—IEF gels were run using ampholytes ranging from 4 to 6 (Bioytes/6; Bio-Rad) as previously described (8). IEF protein markers were purchased from Invitrogen (Carlsbad, CA).

**Peptide Sequencing**—Automated gas-phase amino acid sequencing was performed on an Applied Biosystems (Foster City, CA) model 491 Procise protein sequencer incorporating a model 140C microgradient system and a 785A programmable absorbance detector.

**Characterization of Lipooligosaccharide Cores**—Whole cells of C. jejuni strains were subjected to proteinase K digestion, electrophoresed on 16% Tricine gels (Invitrogen), and stained with silver (Bio-Rad) as described previously (28).

**RESULTS**

**Analysis of Intact Campylobacter Flagellins**—Electrospray mass spectrometry experiments on purified flagellin from C. jejuni 81–176 (Fig. 1) indicated that the molecular mass of the monomeric glycoprotein was ~10% higher than that predicted from the sequence of flaA, 59,240 Da (GenBank accession number AF345999). The reconstructed molecular mass profile obtained from the multiply protonated ions observed in the electrospray mass spectrometry (Fig. 1, inset) showed two components at 65,766 and 65,841 Da with a broad peak profile of

**FIG. 1.** Electrospray mass spectrum of intact flagellin from C. jejuni 81–176. The reconstructed molecular mass profile, shown as an inset, indicates two major peaks at 65,766 and 65,841 Da. The observed mass is ~6.5 kDa higher than the predicted protein (59,240 Da).
Glycosylation of Campylobacter Flagellin

Fig. 2. Capillary LC-ESMS analysis of tryptic peptides from C. jejuni 81–176 flagellin. A, total ion chromatogram acquired at normal orifice voltage of 30 V (solid line, m/z 400–1800) and reconstructed ion chromatograms acquired under front end collision-induced dissociation conditions (OR 120 V) for the unique glycosyl markers ions m/z 317 (dotted line) and m/z 409 (dashed line). Only the tryptic peptides bearing the marker ions m/z 317 and m/z 409 are observed in the reconstructed ion chromatograms. B, extracted mass spectra for the peak at 22.2 min. The multiply charged ions (circled) correspond to the tryptic peptide T390 bearing 5 modifications. The inset shows the molecular mass profile of this peptide. C, extracted mass spectra for the peak at 22.9 min. The multiply charged ions (circled) correspond to the tryptic peptide T200–222 bearing a single modification. The inset shows the molecular mass profile of this peptide.

600–700 Da, possibly reflecting the heterogeneity in the glycoform distribution. It is noteworthy that no peak was observed for the predicted, unmodified FlaA flagellin (59,240 Da), suggesting extensive modification on the protein backbone structure. Mass spectral analyses on flagellin purified from strains NCTC 11168 and OH 4384 also gave broad heterogeneous molecular mass envelopes ranging from 65,800 to 66,400 Da with a few discrete glycoform peaks. Similarly, flagellin obtained from C. coli VC167 T2 exhibited a broad molecular mass distribution extending from 64,500 to 65,400 Da, also showing a 6,000-Da mass excess from the predicted sequence (11). In all the intact C. jejuni and C. coli flagellins examined, the extent of post-translational modifications was substantial, and this resulted in a quantitative incorporation of glycosyl moieties imparting approximately a 10% molecular mass excess on the gene product (see below).

cLC-ESMS of Tryptic Peptides from C. jejuni 81–176 Flagellin—To precisely assign the type and location of these post-translational modifications, mass spectral experiments were performed using cLC-ESMS on tryptic peptides derived from purified flagellin. An OR voltage ramp was set over two distinct scanning periods (period 1: 1 s, OR 100 V m/z 150–400; period 2: OR 30 V, m/z 400–1800) to identify characteristic oxonium ions (period 1) together with precursor ions from which these specific carbohydrate fragment ions were derived (period 2). Under these conditions, Neu5Ac-containing glycopeptides typically yield an abundant oxonium ion at m/z 292. Although this residue was originally suspected as an O-linked glycan, the cLC-ESMS with high/low OR stepping could not unambiguously confirm its presence. Rather, the combined cLC-ESMS analyses revealed the elution of tryptic peptides with unusual modifications as reflected by distinct fragment ions at m/z 317 (dotted line) and 409 (dashed line) obtained from the extracted ion chromatograms shown in Fig. 2a.

The occurrence of these fragment ions (period 1) was used as a diagnostic tool to identify which eluting tryptic glycopeptides observed in the total ion chromatogram (period 2, solid line in Fig. 2a) comprised these unusual modifications. These analyses enabled the identification of at least 7 tryptic glycopeptides, some of which carried multiple modified residues. The number and types of substitution were calculated by comparing the mass of the modified tryptic peptides with those of the expected proteolytic fragments. For example, Fig. 2b shows the extracted mass spectrum of the peak eluting at 22.2 min in Fig. 2a. A number of co-eluting components were observed in Fig. 2b including a triply and doubly charged ions at m/z 838.1 and 1256.6 corresponding to the unmodified tryptic peptide T135–157. However, ions observed at m/z 1055.7 and 1407.6 (circled) could not be matched to any expected tryptic peptides and were associated with a peptide molecule of 4218.9 Da. The inset of Fig. 2b shows the reconstructed molecular mass profile obtained from these ions. The observed mass for this component was in good agreement with that calculated for the modified tryptic peptide T300–412 comprising 3 neutral 316 residues and 2 neutral 408 residues (4219.0 Da). It is noteworthy that the isotopic profile also suggests the occurrence of a related molecular species of 1 Da lower (4218.0 Da), which was later assigned as a substituted analog where one of the 316 Da residues was replaced by a 315-Da monosaccharide (see below). Similarly, the extracted mass spectrum taken at 22.9 min (Fig. 2c) shows a single component of 2649.3 Da with triply and doubly protonated ions at m/z 884.2 and 1325.7, respectively (circled). The molecular mass of the modified peptide shown in Fig. 2c corresponded to T^200–222, bearing a single neutral 316 residue. It is noteworthy that this peptide also displayed heterogeneity in the incorporation of a modified residue for which an oxonium ion was observed at m/z 316 (corresponding to a neutral 315-DA residue, data not shown) instead of m/z 317 (corresponding to a 316-DA residue). The nature of the 316-, 315-, and 408-DA substratants was later assigned as pseudaminic acid (Pse5Ac7Ac), its 5-acetamidino analog (Pse5Am7Ac), and 5,7-N-(2,3-dihydroxypropionyl)-Pse (Pse5Pr7Pr), respectively (see below).

MS-MS Analyses of Tryptic Glycopeptides and ε-Elimination Products—HPLC fractions comprising the suspected tryptic glycopeptides were subjected to tandem mass spectrometry analyses to identify structural features that could be assigned to key functional groups of the unusual carbohydrate residues. As an example, the MS-MS spectrum of the [M + 3H]^{3+} ions previously observed at m/z 884.2 in Fig. 2c for tryptic peptide T^200–222 yielded an intense oxonium ion at m/z 317 together with consecutive peptide bond cleavages resulting in b- and y-type ions (32) (Fig. 3a). These ions correspond to fragment ions where the charge is retained on the N and C terminus of the peptide backbone. The latter information was consistent with that predicted for the tryptic peptide T^200–222 and confirmed the previous assignment. To determine the potential empirical formula of the oxonium fragment ion at m/z 317, accurate mass measurements were obtained using the predicted m/z values of y_{2} and b_{4} fragment ions. Upon recalibra-
tion the mass measurement accuracy across the entire range was within 5 ppm of that calculated for individual b- and y-type fragment ions. Accordingly, the mass of the neutral carbohydrate moiety was determined to be 316.122 Da. The precision of this measurement eliminated glycerol phosphatidylinositol as a possible substituent (neutral residue mass 316.0559 Da), although previous investigations have reported the occurrence of a related structure -glycerophosphate as a post-translational modification on Ser residues of Neisseria meningitidis pili (33). Among the different empirical formulas satisfying the defined mass constraints only C₁₃H₂₀O₇N₂ (M_r 316.126) was retained as a plausible candidate. This was also substantiated by the second generation fragment ions of \textit{m/z} 317 formed by collisional-induced dissociation of the tryptic glycopeptide ions in the orifice/skimmer region of the mass spectrometer (Fig. 3ab). The MS-MS spectrum of this unusual carbohydrate residue was characterized by consecutive losses of neutral groups such as water, ketene (CH₂CO) and formic acid (HCOOH). These experiments indicated that the glycosyl moiety was a diamino sugar containing an acid group, two N-acetyl functionalities, and a modified C₂ side chain. Together with the NMR studies performed on the same HPLC fraction (see below) these data were consistent with a pseudaminic acid (Pse5Ac7Ac), an unusual carbohydrate residue previously found in the lipopolysaccharide (LPS) of \textit{P. aeruginosa} (26).

In addition to Pse5Ac7Ac, other related structural modifications were also found in \textit{C. jejuni} flagellin. For example, a second oxonium ion was observed at \textit{m/z} 316 in a number of peptides from the wild type strain of \textit{C. jejuni} –176, including that of glycopeptide T390–412 (Fig. 2b). Second generation fragment ions of \textit{m/z} 316 of this tryptic glycopeptide were achieved using higher quadrupole resolution settings to avoid selection of fragment ion \textit{m/z} 317. The corresponding MS-MS spectrum showed prominent neutral losses of NH₃ and CH₃CH(NH)(NH₂), consistent with the substitution of one of the two acetamido groups of Pse5Ac7Ac by an acetamidino functionality (Fig. 3c). This substitution resulted in a glycosyl moiety 1 Da lower than that of Pse5Ac7Ac and was termed 5-acetamido-7-acetamido-Pse (Pse5Am7Ac). The occurrence of side chain fragment ions corresponding to the C₁–C₉ backbone that are common to both Pse5Ac7Ac and Pse5Am7Ac (\textit{m/z} 134, 180, and 221) suggested...
that the C₅ and not the C₇ acetamido group was substituted. Indeed, if the acetamidino group was located on C7, the backbone fragment ion m/z 180 would have been shifted to m/z 179. Confirmation of this assignment using NMR spectroscopy is presently under way, and results from this investigation will be reported separately. This related structure was observed in tryptic glycopeptides T₂₀₀–₂₂₀, T₃₉₀–₄₁₂, and T₄₂₃–₄₆₆, although some heterogeneity in the incorporation level of this residue was noted in these peptides. An O-acetyl derivative of Pse₅Ac₇Ac₇Ac₈OAc, Pse₅Ac₇Ac₇Ac₈OAc, was also found in the tryptic glycopeptide T₃₉₀–₄₁₂. However, the precise location of this residue could not be established unambiguously due to the presence of other related O-linked sugars on the same peptide. The glycosidic bonds of the tryptic glycopeptides are more labile than peptide bonds, and deglycosylated fragment ions have a structure identical to that of unglycosylated peptide ions, thus preventing the identification of the modification sites.

Another common modification encountered in C. jejuni 81–176 flagellin was noted previously in Fig. 2b and corresponded to a neutral residue of an accurate mass measurement of 408.139 ± 0.003 Da. The mass difference between this residue and Pse₅Ac₇Ac is 92.027 Da, in good agreement with an incremental C₂H₄O₄ moiety. The second generation product ion from m/z 409 (data not shown) was consistent with a substituted Pse₅Pr₇Pr, whereby the 2 N-acetyl groups were replaced by 2 N-2,3-dihydroxypropionyl groups (Mr 408.1368) (Pse₅Pr₇Pr).

Although the modified peptides were identified based on mass differences from the predicted sequence, the individual amino acids bearing the modifications were not as readily assigned. The labile nature of the glycosidic bond between the hydroxyl amino acid Ser/Thr and the carbohydrate residue made it difficult to observe fragment ions comprising the intact modification. To unambiguously assign the site of O-linked attachment, purified glycopeptide fractions were subjected to base-catalyzed hydrolysis in the presence of NH₄OH whereby the NH₄⁻elimination product incorporated a newly formed amino group of a distinct mass (34). For example, upon NH₄⁻elimination,

---

**Table I**

Assignment of glycosyl substituents observed on tryptic peptides from C. jejuni 81–176 flagellin

| Peptides    | M (Da) | Modification                      | Sequence and location of modification |
|-------------|--------|----------------------------------|---------------------------------------|
| T200-222    | 2648.4 | Pse5Ac7Acₐ, Pse5Am7Acₐ          | FDNVVTSTVGTLGALAEINK                   |
| T338-365    | 3498.6 | Pse5Ac7Acₐ, Pse5Am7Ac₁          | DINSGTNLSAIGMGTTD*MSQSVSLR*           |
| T390-412    | 4218.0 | Pse5Ac7Ac₂₃, Pse5Am7Ac₂₃, Pse₅Pr₇Pr₂ | FVHTONVSSISAQMSPGSGFSR*               |
| T390-422    | 5441.0 | Pse5Ac7Ac₂₃, Pse₅Ac7AcesOAc₈, Pse₅Pr₇Pr₂ | FVHTONVSSISAQMSPGSGFSR*GSGFSVGSFK*   |
| T413-422    | 1239.4 | Pse5Ac7AcesOAc₁                 | GSGFSVGSFK*                           |
| T423-466    | 6898.8 | Pse5Ac7Acₐ, Pse5Am7Ac₁, Pse₅Pr₇Pr₁ | NLNSGLSQIQNSSAASMSNTTVVSGSFSGSSGSONQFAALK* |
| T467-483    | 2009.8 | Pse5Ac7Ac₁                     | TTAANTTD*ETAGVITLK*‡‡‡                 |

a Corresponds to monoisotopic molecular mass.

b The single underline indicates a modified residue, whereas a double underline refers to the position of a Pse5Pr₇Pr substituent.

c Assignment is based on a previous report of modified residues on C. coli VC167 tryptic peptide by Edman sequencing (9).

d Also confirmed by Edman sequencing of 81–176 tryptic peptide.

* The peptide sequence was confirmed by DNA sequencing of the polymerase chain reaction product corresponding to base pairs 1–1432 (GenBank™ accession number AF345999).
O-linked Ser and Thr residues yield modified amino acids of neutral mass of 86 and 100 Da, respectively. This is illustrated in Fig. 4 for the tryptic glycopeptide previously shown in Figs. 2c and 3e whereby a total of four potential O-linked sites are present in the proteolytic fragment T200–222. In this particular case, only Ser206 was modified with a Pse5Ac?Ac residue, as evidenced by a mass shift of 86 Da between y16 and y17 (Fig. 4). By using this technique 10 of the 19 modifications sites found on C. jejuni flagellin could be uniquely identified to their corresponding Ser or Thr residues as is seen in Table I. An additional nine sites have been tentatively assigned in peptide T466–486 based on Edman sequencing of the corresponding peptide from C. coli VC167 flagellin (9). In some cases, Pse5Ac7Ac analogues were assigned to specific Ser residues due to their relative lability compared with other occupied sites. This was the case for tryptic glycopeptide T390–412, where Pse5Pr7Pr residues were assigned to Ser397 and Ser404, whereas Thr393, Ser400, and Ser408 were modified with Pse5Ac7Ac residues. Such a situation was found to be the exception rather than the rule since our MS-MS analyses mainly indicated types of modifications found on individual peptides rather than assigning the exact substituent attached to each Ser or Thr residue. Table I summarizes these findings and indicates the location and modifications found along the flagellin protein backbone. By using a combination of LC-ESMS and MS-MS experiments, 19 sites of modifications were identified on C. jejuni 81–176 flagellin. The predominant modification was Pse5Ac7Ac, along with the related structures Pse5Am7Ac, Pse5Ac7Ac8OAc, Pse5Pr7Pr. The mass excess associated with these substitutions (ΔM, 6411 Da) was consistent with that measured on the intact flagellins (~6.5 kDa) and with a high level of site occupancy (between 18–20 sites, Fig. 1). Although some microheterogeneity was observed on the different O-linked Ser/Thr residues, these experiments indicate that all 19 identified sites are usually occupied in each flagellin monomer.

**NMR Analysis**—Mass spectral analyses of the modifications on both C. jejuni 81–176 and C. coli VC167 flagellin had indicated that the predominant glycosyl moiety was Pse5Ac7Ac. Tryptic peptide T200–222 was selected for NMR analysis to unambiguously identify Pse5Ac7Ac. However, since C. jejuni 81–176 tryptic peptide T200–222 contained either Pse5Ac7Ac or Pse5Am7Ac at Ser206 (see Table I), the C. coli VC167 tryptic peptide T200–222, which was modified with only Pse5Ac7Ac, was selected for NMR analysis. Either the 1-glycero-1-manno (pseudaminic acid), the 1-glycero-3-manno, or the 1-glycero-3-galacto, or the 1-glycero-3-talo configuration has been found in bacterial LPS (35–38). From the spectra of the tryptic peptide T200–222, only resonances characteristic for pseudaminic acid were clearly observed among the peptide resonances (Fig. 5). In the COSY spectrum, along with the (H-3ax, H-3eq) cross-peak at (1.62, 2.06) ppm (not shown), the cross-peaks to (H-4, H-3ax) and (H-4, H-3eq) were observed at (4.07, 1.62) and (4.07, 2.06) ppm in Fig. 5A. The (H-4, H-5) and (H-5, H-6) cross-peaks were also clearly observed at (4.07, 4.22) and (4.22, 3.81) ppm in the COSY spectrum (Fig. 5B). An intense (H-6, H-7) cross-peak was observed in the NOE spectroscopy plot in panel C, the cross-peaks of protons that are in close proximity to the CH2 (H-3ax and H-3eq), CH3 (H-9), and NAc protons are shown. In panel D, the NOE cross-peaks for the other sugar resonances are shown.
also observed at (3.81, 4.17) ppm typical of the large J6,7 proton coupling constant of 10 Hz found only for pseudaminic acid and not the other isomers. The isolated resonance at 1.18 ppm, characteristic of a CH₃ group, with a proton vicinal coupling constant of 6.6 ± 0.2 Hz was assigned as H-9. From the COSY spectrum in Fig. 5A, the H-8 resonance was located at 4.17 ppm. The H-8 and H-7 resonances were overlapping since a small (H-9, H-6) cross-peak was also observed at (1.18, 3.81) ppm. The two singlets at 1.97 and 1.99 ppm have characteristic chemical shifts of NAc groups. Hence, all the expected proton chemical shifts similar to those found for 5,7-diacetamido-3,5,7,9-tetrahydroxy-L-glycero-D-manno-nonulosonic acid (26, 27) were unambiguously located.

The observed NOEs (Fig. 5, C–D) were also found to be consistent with the interproton distances obtained from a molecular model of 5,7-diacetamido-3,5,7,9-tetrahydroxy-L-glycero-D-manno-nonulosonic acid as drawn in Fig. 5. Since NOEs are highly dependent on interproton distances (r⁻⁶), they are thus also dependent on the structure of the molecule and the correct assignments. A strong NOE (not shown) was observed between the H-3ax and H-3eq resonances, in accord with an interproton distance of 2.5 Å. In Fig. 5C, the (H-4, H-3ax) NOE was bigger than the (H-4, H-3eq) NOE, in accord with respective interproton distances of 2.5 and 3.1 Å. Strong NOEs were also observed for (H-8, H-9) and (H-6, H-9), consistent with interproton distances of closest approach of 2.5 and 2.3 Å, respectively. In Fig. 5D, the observed NOEs between (H-4, H-5) and (H-5, H-6) and (H-4, H-6) were also consistent with their interproton distance of 2.4, 2.5, and 2.5 Å, respectively. In Fig. 5C, small NOEs between the NAc resonances at 1.97 and 1.99 ppm and the H-5 and H-7 resonances were consistent with interproton distances for (NAc-5, H-7) of 3.6 Å, (NAc-5, H-5) of 4.5 Å, (NAc-7, H-5) of 3.7 Å, and (NAc-7, H-7) of 4.5 Å. Hence, all the NMR data was in accord with the glycoside moiety being a pseudaminic acid.

Characterization of Genes Involved in Flagellin Modification—Linton et al. (12) report that mutation of gene Cj1317, annotated as neuB3 encoding a sialic acid synthase in the genome sequence of C. jejuni NCTC 11168, resulted in a non-flagellated phenotype in several strains of C. jejuni. The corresponding region of the genome of C. jejuni 81–176 was cloned. Sequence analysis revealed orthologs of Cj1317, Cj1316c, Cj1315c, and Cj1314c in an order similar to that described in the genome sequence of C. jejuni strain NCTC 11168 (see Table II). Insertional inactivation of the neuB3 gene (Cj1317) with a chloramphenicol cassette (Cm') resulted in loss of motility, and no flagellin was detected in whole cell lysates by Western blotting with a rabbit polyclonal antiserum against 81–176 flagellin (data not shown), confirming the results of Linton et al. (12). However, mutation of the adjacent gene, Cj1316c, encoding a predicted protein of 43.7 kDa in C. jejuni NCTC 11168 (22), resulted in a motile phenotype. The predicted protein encoded by this gene also shares significant similarity to the product of Cj1324c as well as to two genes involved in LPS biosynthesis in Legionella pneumophila and P. aeruginosa O5 serogroup, involved in synthesis of legionaminic acid and mannuronic acid, respectively (see Table II). Flagellin from the Cj1316c mutant was compared with flagellin from wild type 81–176 on IEF gels. Flagellin from wild type 81–176 separated into multiple glycoforms ranging from approximate pl 4.2–5.2 (Fig. 6B, lane 1), similar to other Campylobacter flagellins (11, 28). In comparison, flagellin from the Cj1316c mutant also showed multiple glycoforms but at a pl range of ~3.5–4.4 (Fig. 6B, lane 2). This shift toward the more acidic

### Table II

| 81–176 gene | % identity/similarity (length in aa) | Homologue | Annotation (accession No.) |
|-------------|-------------------------------------|-----------|---------------------------|
| neuB3       | 92/93 (343)                         | C. jejuni NCTC 11168 CJ1317 | NeuNac synthetase/flagellar (B81275) |
|             | 41/55 (329)                         | Aeromonas NeuB | LPS biosynthesis/flagellar modification (AAD45660) |
| pseA        | 40/56 (329)                         | Caulobacter NeuB | Putative NeuNac synthetase (AAK24832) |
|             | 99/99 (378)                         | C. jejuni NCTC 11168 CJ1316c | Unknown (A81275) |
|             | 33/54 (373)                         | C. jejuni NCTC 11168 CJ1324c | Unknown (A81275) |
|             | 30/48 (373)                         | L. pneumophila Orf28 | LPS biosynthesis (CAB65216) |
|             | 32/53 (119)                         | P. aeruginosa WbpG | B band LPS biosynthesis (E83253) |
| Cj1315c     | 99/99 (201)                         | C. jejuni NCTC 11168 CJ1315c HisH1 | Amidotransferase/histidine synthesis (Q9PFM4) |
|             | 46/65 (201)                         | Methanococcus jannaschi HisH | Histidine biosynthesis (Q9Z329) |
|             | 42/67 (202)                         | P. aeruginosa HisH2 | Putative cyclase (G81274) |
| Cj1314c     | 93/93 (248)                         | C. jejuni NCTC 11168 CJ1314c | Histidine biosynthesis (P72139) |
|             | 42/67 (248)                         | P. aeruginosa HisF2 | Histidine biosynthesis (P72139) |

---

**Fig. 6. A, Schematic of relevant C. jejuni 81–176 genes. Arrows indicate the location and orientation of the Cm’ cassette. The insertion into Cj1316c was 661 base pairs from the translational start; the insertions into Cj1316c, Cj1315c, and Cj1314c were 774, 199, and 126 base pairs, respectively, into the genes. The region of the 81–176 chromosome cloned in pSG1854 and subcloned into pRY107 is shown. The inset in pSG1854 begins at the end of Cj1317 and extends 618 base pairs 3’ to the end of Cj1314c. B, IEF of flagellins. Flagellins were separated in an ampholyte mixture of pH 4–8 as previously described (8, 25). Lane 1, 81–176; lane 2, 81–176 pseA (Cj1316c) mutant; lane 3, 81–176 pseA (pRY107/1854); lane 4, 81–176 (pRY107). kb, kilobase pair.**
region of the IEF gel is consistent with the loss of a basic functionality on the Pse5Ac7Ac structure. The two genes downstream of Cj1316c, Cj1315c, and Cj1314c encode proteins annotated as homologs of HisH and HisF in C. jejuni NCTC 11168 (22), respectively (see Table II). Mutations in Cj1315c and Cj1314c resulted in a motile phenotype, and flagellins isolated from both mutants displayed IEF patterns identical to wild type 81–176 (data not shown). Thus, the insertion into Cj1316c did not appear to exert a polar effect on the downstream genes. Nonetheless, to confirm that the flagellin phenotype observed was due to mutation at this locus and not to phase variation (12, 22) at a distant site, the Cj1316c mutant was complemented in trans with a fragment of DNA containing Cj1316c through Cj1314c in plasmid pRY107/1854 (see Fig. 6B). Flagellin isolated from the complemented mutant (Fig. 6B, lane 3) displayed an IEF pattern intermediate between wild type flagellin (lane 1) and that from the Cj1316c mutant (Fig. 6B, lane 2). This partial complementation is likely due to the presence of the genes on a multicopy plasmid. Flagellin from the Cj1316c mutant containing the pRY107 shuttlle plasmid alone appeared identical to the Cj1316c mutant (Fig. 6B, lane 4). The mobility of the lipooligosaccharide cores of wild type 81–176 and those of all four mutants were compared on 16% Tricine gels, and there were no differences observed under conditions that have been shown to detect the loss of a single sugar residue from lipooligosaccharide cores (Ref. 28; data not shown).

LC-MS hi/low orifice stepping with reconstructed ion chromatograms on m/z 316 and m/z 317 was done to compare flagellins from the 81–176 parent, the Cj1316 mutant, and the mutant complemented in trans with pRY107/1854. Flagellin from both 81–176 and the Cj1316c mutant carrying pRY107/1854 contained the m/z 316 fragment ion characteristic of Pse5Am7Ac (data not shown). In contrast, flagellin from the Cj1316c mutant no longer displayed this modification, as evidenced by the absence of the relevant signal in the m/z 316 channel for the expected tryptic peptides. Instead all sites previously occupied by Pse5Am7Ac in the wild type flagellin were replaced by Pse5Ac7Ac residues in the Cj1316c mutant. Based on these data, we propose that Cj1316c be named pseA (pseudaminic acid), indicating its role in the biosynthetic pathway of this related group of molecules.

**DISCUSSION**

The results presented here show that *Campylobacter* flagellin is one of the most extensively modified prokaryotic proteins identified to date. A total of 19 sites of O-linked modification have been characterized on flagellin from strain 81–176, representing 10% of the total mass of the protein. The extent of the modifications accounts for the discrepancy observed between predicted mass and the accurate mass measurement obtained in this study. A similar degree of modification was found on the Cj1316c mutant, indicating that Pse5Am7Ac and derivatives are responsible for the discrepancy observed between predicted mass and the accurate mass measurement obtained in this study. A similar degree of modification was found on the Cj1316c mutant, indicating that Pse5Am7Ac and derivatives are responsible for the discrepancy observed between predicted mass and the accurate mass measurement obtained in this study.

Two lines of evidence suggest that Neu5Ac was a component of the posttranslational modifications of *Campylobacter* flagellin (7, 11, 13). Doig et al. (8) show that the sialic acid-specific lectin limax flavus agglutinin bound to *Campylobacter* flagellins, but this observation was likely due to cross-reaction of the lectin to the structurally similar Pse5Ac7Ac moieties. Additionally, genomic sequencing of *C. jejuni* NCTC 11168 revealed the presence of multiple alleles of genes encoding proteins predicted to be involved in Neu5Ac biosynthesis (22). One set of these genes is clearly involved in biosynthesis of the Neu5Ac found in the lipooligosaccharide core of *C. jejuni* (11, 28), and genetic evidence has suggested that the other two sets of putative Neu5Ac genes are involved in flagellin modifications. Thus, mutation in a gene encoding a putative CMP-Neu5Ac synthetase, termed *pmtB* (11), and another in a putative Neu5Ac synthase, *neuB3* or Cj1317, have been shown to affect flagellin (Refs. 11 and 13 and this work). The E. coli K1 Neu5Ac synthase, NeuB, condenses mannosamine and phosphoenolpyruvate to form Neu5Ac. The homology of the *neuB2* (Cj1327c) and *neuB3* gene products to *E. coli* K1 NeuB suggests that the enzyme encoded by either of these genes may be involved in the condensation of a C6 and a C3 sugar precursor to form Pse5Ac7Ac in a similar fashion (42).

The definition of the chemical structure of the glycosyl modifications on flagellin will facilitate elucidation of the enzymatic pathways. Thus, in this study we have identified a role of the Cj1316c gene product or PseA in the synthesis of Pse5Am7Ac. As seen in Table II, PseA shows homology to proteins involved in synthesis of two related structures, legionaminic acid and mannnuric acid in *L. pneumophila* and *P. aeruginosa*, respectively. Both of these structures contain molecules with an acetamidino functionality (43, 44). Interestingly, both the *L. pneumophila* and *P. aeruginosa* LPS gene clusters also contain orthologs of hisH and hisF, to which Cj1315c and Cj1314c show homology (see Table II). Thus, we cannot exclude a role for these genes in flagellin glycosylation since our preliminary mutant screening required detection of change in IEF gels. Additional experimentation will be required to determine if Cj1315c and Cj1316c are involved in flagellin glycosylation.

The basis of O-linked glycosylation is poorly defined. Only selected serines in *Campylobacter* flagellin were decorated with a glycosyl group, and Pse5Pr7Pr, Pse5Ac7Ace8OAc, or Pse5Am7Ac are constrained to certain residues, suggesting specificity in the glycosylation process. However, it is interesting that in the absence of Pse5Am7Ac in the pseA mutant, the corresponding residues were replaced with Pse5Ac7Ac. Most of the modified residues were observed in a narrow hydrophobic region of the central core domain located between residues 342 and 481, a region shown to be surface-exposed in the flagella filament (10). The corresponding region in flagellin from *C. coli* VC167 had been shown to contain 12 modified serines (9), all of which are conserved in 81–176 flagellin. Based on the glycosylated residues identified in both 81–176 and VC167 flagellins (9), the site of attachment does not appear related to a consensus peptide sequence. Rather, the site of glycosylation appears to be at least partially dependent on local hydrophobicity upstream of Ser/Thr residues (boxed residues in Table I). Hydroxylated residues lying downstream from this local hydrophobic environment are expected to project outward, thereby rendering them accessible to glycosyl transferases (see Table I). Ser and Thr adjacent to acidic or basic residues (see asterisks, Table I) in the central peptide region were not typically glycosylated (e.g. Ser257, Thr354, Thr473, Thr476). O-Linked glycosylation in *C. jejuni* flagellin may be a partially selective process preferentially targeting residues surrounded by aliphatic/aromatic residues. This is consistent with the surface-exposed site of the single glycolyl moiety added to *N. gonorrhoeae* pilin at Ser243 (45) but distinct from processes proposed for other prokaryotic O-linked glycosylated proteins (15, 16).

The function of glycosylation to *Campylobacter* flagellin remains to be determined. The modifications will increase the hydrophilicity of flagellin and are likely to influence the inter-
actions of C. jejuni with eukaryotic cells. The structural similarity of Pse5Ac7Ac to Neu5Ac might also suggest a role in immune avoidance, although there is evidence that at least some of the modifications are immunogenic and antigenically variable within a strain (8–11). Moreover, the non-motile phenotype of neuB3 mutants (Ref. 12 and this work) suggests that glycosylation may affect Campylobacter flagellin subunit interactions and/or assembly, just as it does for Halobacterium flagellin (46) and Neisseria pilin (47).

In this report we have used mass and NMR spectroscopy to characterize the structure and extent of glycosylation on Campylobacter flagellin. The identification of the genes involved in synthesis and addition of the glycosyl moieties to flagellin can now be addressed. The data obtained will form the basis for studies on the role of these glycosyl moieties in both flagellar structure and function as well as C. jejuni virulence. Moreover, the system provides an excellent model system to study the process of O-linked glycosylation in prokaryotes.

Acknowledgments—We thank Dawn Pattarini for purification of flagellin, Dr. Y. Knirel for the preprint of his review on pseudaminic acid in bacterial glycopolymers, Dr. N. M. Young and Dr. E. Altman for flagellin, Dr. Y. Knirel for the preprint of his review on pseudaminic acids and/or assembly, just as it does for Halobacterium flagellin can now be addressed. The data obtained will form the basis for studies on the role of these glycosyl moieties in both flagellar structure and function as well as C. jejuni virulence. Moreover, the system provides an excellent model system to study the process of O-linked glycosylation in prokaryotes.

REFERENCES
1. Friedman, C. R., Neimann, J., Wegener, H. C., and Tauxe, R. V. (2000) J. Infect. Dis. 180, 5010–5019
2. Brimer, C. D., and Monie, T. C. (1998) J. Bacteriol. 180, 3209–3217
3. Kurlath, J. A., Osterholm, M. T., Judy, A., Forgung, J. C., and Robinson, R. A. (1985) J. Infect. Dis. 152, 592–596
4. Parkhill, J., Wren, B. W., Mungall, K., Kettle, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V., Meade, S., Pallen, M. J., Penn, C. W., Quail, M. A., Rajandream, M. A., Rutherford K. M., van Vliet, A. H., Whitehead, S., and Barrell, B. G. (2000) Nature 403, 665–668
5. Gilbert, M., Brison, J. R., Kawarski, M. F., Michniewicz, J., Cunningham, A. M., Wo, Y., Young, N. M., and Wakschul, W. (2000) J. Biol. Chem. 275, 3896–3906
6. Manzi, A. E., and Keifer, P. A. (1997) in Techniques in Glycobiology (Townsend, R. R., and Hotchkiss, A. T., eds) 1–16, Marcel Dekker, Inc., New York
7. Uhrin, D., and Brison, J. R. (2000) in NMR in Microbiology: Theory and Applications (Barbotin, J. N., and Portais, J. C., eds) pp. 165–210, Horizon Scientific Press, Wymondham, UK
8. Knirel, Y. A., Koczarowa, N. A., Shashkov, A. S., Dmitriev, B. A., Kochechkov, N. K., Stanislavsky, E. S., Maskell, D. J., and Mashilova, G. M. (1987) Eur. J. Biochem. 163, 639–652
9. Zenne, L., Lindberg, B., Schweda, E., Gustafsson, B., and Holme, T. (1988) Carbohydr. Res. 180, 285–294
10. Guerry, P., Ewing, C. P., Hickey, T. E., Prendergast, M. M., and Moran, A. P. (2000) Infect. Immun. 68, 6656–6662
11. Guerry, P., Yao, R., Alm, R. A., Barr, D. H., and Trust, T. J. (1994) Methods Enzymol. 235, 474–481
12. Yao, R., Alm, R. A., Trust, T. J., and Guerry, P. (1993) Gene 138, 258–268
13. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J, and Guerry, P. (1999) J. Biol. Chem. 274, 38796–38806
14. Acknowledgments—We thank Dawn Pattarini for purification of flagellin, Dr. Y. Knirel for the preprint of his review on pseudaminic acid in bacterial glycopolymers, Dr. N. M. Young and Dr. E. Altman for valuable discussions, Dr. M. Gilbert for strain OH4384, D. Watson for Edman sequencing analyses, and T. Devescovi for photographic assistance.

Glycosylation of Campylobacter Flagellin
Identification of the Carbohydrate Moieties and Glycosylation Motifs in Campylobacter jejuni Flagellin

Pierre Thibault, Susan M. Logan, John F. Kelly, Jean-Robert Brisson, Cheryl P. Ewing, Trevor J. Trust and Patricia Guerry

J. Biol. Chem. 2001, 276:34862-34870.
doi: 10.1074/jbc.M104529200 originally published online July 18, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M104529200

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 43 references, 18 of which can be accessed free at http://www.jbc.org/content/276/37/34862.full.html#ref-list-1