Structure of the N-Linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, *Campylobacter jejuni*

Received for publication, June 19, 2002, and in revised form, August 7, 2002
Published, JBC Papers in Press, August 16, 2002, DOI 10.1074/jbc.M206114200

N. Martin Young‡, Jean-Robert Brisson, John Kelly, David C. Watson, Luc Tessier, Patricia H. Lanthier, Harold C. Jarrell, Nicolas Cadotte, Frank St. Michael, Erika Aberg, and Christine M. Szymanski§

From the Institute for Biological Sciences, National Research Council of Canada, 100 Sussex Dr., Ottawa, Ontario K1A 0R6, Canada

Mass spectrometry investigations of partially purified *Campylobacter jejuni* protein PEB3 showed it to be partially modified with an Asn-linked glycan with a mass of 1406 Da and composed of one hexose, five N-acetylhexosamines and a species of mass 228 Da, consistent with a trideoxydiacetamidohexose. By means of soybean lectin affinity chromatography, a mixture of glycoproteins was obtained from a glycan extract, and two-dimensional gel proteomics analysis led to the identification of at least 22 glycoproteins, predominantly annotated as periplasmic proteins. Glycopeptides were prepared from the glycoprotein mixture by Pronase digestion and gel filtration. The structure of the glycan was determined by using nano-NMR techniques to be GalNAc-α1,4-GalNAc-α1,4-[Gluβ1,3]-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bacβ1,β1,N-Asc-Xaa, where Bac is bacillosamine, 2,4-diacetamido-2,4,6-trideoxyglycrophoramide. Protein glycosylation was abolished when the *pglB* gene was mutated, providing further evidence that the enzyme encoded by this gene is responsible for formation of the glycopeptide N-linkage. Comparison of the *pgl* locus with that of *Neisseria meningitidis* suggested that most of the homologous genes are probably involved in the biosynthesis of bacillosamine.

Glycosylation of proteins was once considered to be specifically a eukaryotic phenomenon, but it is now clear that it is widespread in both the Archaea and Eubacteria domains (1, 2). Glycosidic linkages of both the N- and O-types have been identified in a diverse group of prokaryotic organisms with a preponderance of N-linked sugars apparent in the Archaea, whereas linkage units of the O-type predominate in glycoproteins identified thus far in the Eubacteria (1, 2). In addition, bacterial N- and O-linkages are formed with a wider range of sugars than those observed in eukaryotic glycoproteins.

Recently a gene locus was identified in the enteric pathogen *Campylobacter jejuni*, which appears to be involved in the glycosylation of multiple proteins and which provided the first evidence of a pathway for widespread protein glycosylation in a Gram-negative bacterium (3). Mutagenesis of genes within this locus, termed *pgl* (for protein glycosylation), resulted in loss of immunogenicity in multiple proteins. The glycan moieties of these proteins were also shown to be recognized by antisera from experimentally infected human volunteers (3). Removal of the glycan moieties by *pgl* mutation resulted in decreased adherence and invasion in vitro and loss of mouse colonization in vivo (4), suggesting that protein glycosylation influences the virulence properties of this organism. A similar genetic locus has recently been described for *pilin* O-glycosylation in *Neisseria meningitidis* (5, 6).

This glycosylation pathway appears to be distinct from that involved in the glycosylation of flagellin observed for several bacterial species (7–10). Thibault et al. (11) recently demonstrated that the glycosyl moiety on the flagellin of *C. jejuni* consists of variants of an unusual nine-carbon sugar, pseudaminic acid, O-linked to Ser or Thr. This sugar may be considered a highly modified sialic acid (12).

We have been identifying and characterizing post-translational modifications of proteins in *C. jejuni* strain NCTC 11168, the strain for which the whole genome sequence has been described by Parkhill et al. (13). Among the proteins giving rise to multiple spots on two-dimensional gels was PEB3, or *Cj*0289c, a major antigenic protein of *C. jejuni* first described by Pei et al. (14). When purified and analyzed by one-dimensional SDS-PAGE, it revealed two bands with a mass difference of ~1500 Da, both of which had N-terminal sequences corresponding to authentic PEB3. Concurrent with our observations on PEB3, Linton et al. (15) identified two putative glycoproteins from *C. jejuni* by use of the GalNAc-specific lectin, soybean agglutinin, one of which was PEB3, and the other a putative periplasmic protein *Cj*1670c, which they named *CgpA*. The authors also observed a number of other putative glycoproteins, based upon their ability to bind to the lectin, but these were not identified. Furthermore, protein binding to the lectin was also affected by mutagenesis of genes in the *pgl* locus.

We have undertaken a comprehensive analysis of the *C. jejuni* NCTC 11168 glycoprotein complement, and in this report we present data on the identification of individual glycoproteins and structural characterization of the N-linked glycan moiety, which is common to all these glycoproteins, by mass spectrometry and NMR spectroscopy. In addition, mutation of a gene, *pglB*, whose homology to the STT3 subunit of the N-linked oligosaccharyltransferase of *Saccharomyces cerevisiae* suggested a role in glycoprotein biosynthesis (3, 16), was shown to specifically affect the glycosylation of the identified glycoproteins.

**EXPERIMENTAL PROCEDURES**

**Bacterial Strains and Plasmids—** *C. jejuni* NCTC 11168 was routinely grown on Mueller Hinton agar under microaerophilic conditions (10% CO₂, 5% O₂, 85% N₂) at 37 °C. *Escherichia coli* DH10B (Invitrogen) was used as the host strain for cloning experiments, and clones were grown on S-Gal™/LB agar (Sigma) or MH agar at 37 °C. When appropriate, antibiotics were added to the following final concentra-
Biosciences) was used on an Applied Biosystems Inc. sequencing system. The PEB3 protein was purified to homogeneity by cation exchange chromatography of the glycan extract as previously described. A Mono S HR 5/5 column (Amersham Biosciences) was used on an ÄKTA Explorer LC system (Amersham Biosciences). The column eluate was monitored for UV absorbance at 280 nm, and fractions were examined by SDS-PAGE analysis in Mini Protein II slab gels (Bio-Rad Laboratories). N-terminal sequencing of individual proteins was performed on a model 491 Procise protein sequencer. The extracts were clarified by centrifugation at 10,000 \( \times g \) for 15 min and immediately frozen at \(-75^\circ C\). Frozen cell pellets were thawed on ice in 0.2 M glycine HCl buffer, pH 2.2 (17), and extracted for 15 min with gentle stirring. Extracts were clarified by centrifugation at 10,000 \( \times g \) for 15 min, dialyzed against pure water (Milli-Q system, Millipore Corp.), and freeze-dried.

**Purification and Analysis of PEB3** — The PEB3 protein was purified to homogeneity by cation exchange chromatography of the glycan extract as previously described (14). A Mono S HR 5/5 (Amersham Biosciences) was used on an ÄKTA Explorer LC system (Amersham Biosciences). The column eluate was monitored for UV absorbance at 280 nm, and fractions were examined by SDS-PAGE analysis (18) in Mini Protein II slab gels (Bio-Rad Laboratories). N-terminal sequencing of individual proteins was performed on a model 491 Procise protein sequencer (Applied Biosystems Inc.), following transfer from the sequencing system (Applied Biosystems Inc.) to the manufacturer’s instructions and resolved by isoelectric focusing on the precast IEF strips followed by SDS-PAGE on homogeneous 12.5% slab gels, 20 \( \times \) 20 cm, for the second dimension. Gels were stained with Bio-Rad’s Coomassie Blue stain (Bio-Rad Laboratories) or silver-stained (20). For subsequent lectin probing, the gels were electroblotted onto PVDF membrane at 50 V for 1 h in 10 mM 3-(cyclohexylamino)-1-propanesulfonic acid buffer, pH 11, containing 10% acetonitrile. The membrane was washed in Milli-Q water and blocked in Tris-buffered saline (100 mM NaCl, 50 mM Tris, pH 7.5) with 0.05% Tween 20 and 2% blocking buffer (Roche Molecular Biochemicals) for 2 h at room temperature. Following blockage, the gels were further incubated with SBA-alkaline phosphatase conjugate (EY Laboratories Inc.) at a concentration of 10 \( \mu g/ml \) in the above blocking solution for 1 h at room temperature. Blots were washed three times in Tris-saline with 0.05% Tween 20 and developed using nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate in 0.1 M NaCl, 0.1 M Tris, pH 9.5, was used as substrate.

**Analysis of Tryptic Glycopeptides** — Glycopeptides were analyzed by MALDI-TOF MS using a MaldiTOF LR mass spectrometer (Micromass). Approximately 0.5 \( \mu l \) of the MALDI matrix solution (10 mg/ml a-cyano-4-hydroxy-cinnamic acid in 50% acetonitrile, 0.2% trifluoroacetic acid) was deposited on the target plate and allowed to dry. The peptide extracts were desalted using C18 ZipTips (Millipore) and were deposited directly on the matrix spots. Acquisition of the MALDI-TOF MS spectra was carried out automatically. The spectra were calibrated externally using peptide standards and internally with trypsin autolysis. Data base searching was carried out in batch mode using Mascot Daemon (Matrix Science) and against the C. jejuni NCTCS P11168 genome sequence data base.

The extracts from the fainter protein spots were analyzed by NanoLC-MS/MS using the Q-TOF2 mass spectrometer. The entire samples were injected onto a 0.3 \( \times \) 5-mm C18 micro pre-column cartridge (Thermo Electron) and then onto a 75-\( \mu \)m C18 Nano-Series column (Dionex/LC-Packings) and resolved by gradient elution (5–90% acetonitrile, 0.2% formic acid in 45 min). The mass spectrometer was set to operate in automatic MS/MS acquisition mode, and spectra were acquired on doubly, triply, and quadruply charged ions. The protein spots were excised and destained with a 1:1 ratio of 30 mM potassium ferricyanide and 100 mM sodium thiosulfate (21). The gel spots were washed extensively with de-ionized water, shrunk with acetonitrile, and re-swollen with 50 mM ammonium bicarbonate containing Promega modified trypsin (10 ng/\( \mu l \)), and sufficient 50 mM ammonium bicarbonate was added to cover the gel pieces (typically 30 \( \mu l \)). The tubes were sealed and incubated overnight at 37 °C. The digest solutions were removed, and the gel pieces were extracted with 50 \( \mu l \) of 5% acetic acid and then with 50 \( \mu l \) of 5% acetic acid in 50% aqueous methanol. The extracts were pooled with the digest solutions and concentrated to \(-10 \mu l \). The peptide extracts from the intense protein spots were analyzed by MALDI-TOF MS using a MaldiTOF LR mass spectrometer (Micromass). Approximately 0.5 \( \mu l \) of the MALDI matrix solution (10 mg/ml a-cyano-4-hydroxy-cinnamic acid in 50% acetonitrile, 0.2% trifluoroacetic acid) was deposited on the target plate and allowed to dry. The peptide extracts were desalted using C18 ZipTips (Millipore) and were deposited directly on the matrix spots. Acquisition of the MALDI-TOF MS spectra was carried out automatically. The spectra were calibrated externally using peptide standards and internally with trypsin autolysis. Data base searching was carried out as described above.

**NMR Spectroscopy** — All spectra were acquired using a Varian Inova 600-MHz spectrometer using the standard Varian software. A gradient 4-mm indirect detection high resolution magic angle spinning nano-NMR probe (Varian) with a broadband decoupling coil was used. The sample was spun at 3 kHz with dry nitrogen as the drive and bearing gas. The spin rate was not computer-controlled but remained constant.
to within 10 Hz of the set value. Samples in 40 μl of D2O solution were recorded at 25 °C and at 35 °C to produce sharper peaks. The pH was unknown due to the small volume. Deuterated EDTA (CDN Isotopes Inc.) was added to chelate metal ions and provide sharper peaks for bacilloside and amino acids. Although the glycopeptide isolate from a P4 column contained some amino acid and sugar impurities, spectra were of sufficient quality to allow complete resonance assignments of the glycopeptide in the presence of 15 mM of deuterated EDTA. Much of the NMR structural work proceeded with this sample, because of the risk of losing the bulk of the isolated glycopeptide by further purification. The derived structure was confirmed by additional NMR experiments on glycopeptide that had been purified using a P2 column, lyophilized, and dissolved in 40 μl of D2O with 1 mM deuterated EDTA. The experiments were performed with suppression of the HOD signal at 4.78 ppm (25 °C) and 4.67 ppm (35 °C). Acquisition and processing of two-dimensional experiments (COSY, TOCSY, NOESY, HMQC, HMBC) were performed as described previously (23). The 1H reference was set by external acetone at 2.23 ppm. The 13C reference was set with the methyl resonance of external acetone at 31.07 ppm. The 1H and 13C chemical shifts in Table I (see below) were measured from the proton spectra and from C-H cross-peaks in the HMQC and HMBC spectra. One-dimensional TOCSY experiments with various spin-lock times from 30 to 151 ms and one-dimensional NOESY with mixing times from 400 to 800 ms were performed as described previously (23, 24). Selective experiments were described as one-dimensional EXP (selected spins, selective excitation bandwidth, mixing time) where EXP is TOCSY or NOESY.

The use of magic angle spinning (MAS) for liquid state samples in the presence of both radio frequency and magnetic-field homogeneities has been shown to influence significantly the performance of mixing sequences in TOCSY experiments and can degrade performance (25, 26). Use of adiabatic (WURST) mixing sequences can eliminate such effects (25, 27). The standard two-dimensional TOCSY and one-dimensional TOCSY sequences were modified so that the MLEV-17 or DIPII-2 mixing sequence was replaced with the adiabatic WURST-2 pulse. The adiabatic (WURST-2) mixing had a single adiabatic inversion pulse length of 1/4 MAS spin rate, a modulation depth of 8, and an adiabaticity of 2. Typically, for the WURST-2 pulse, the sweep bandwidth was 24 kHz, Tm = 0.33 ms (at a MAS spin rate of 3000 ± 10 Hz), B1 (max) = 8.51 kHz, and B1 (root mean square) = 4.77 kHz.

GC-MS Analysis—The enantiomeric configurations of the Glc and GalNAc components of the P2 product were assigned by characterization of the but-2-yl glycosides in gas liquid chromatography-mass spectrometry (28). The derivatives were analyzed using a Hewlett-Packard chromatograph equipped with a 30-m DB-17 capillary column (180–260 °C at 3.5 °C/min), and spectra in the electron impact mode were obtained with a Varian Saturn II mass spectrometer.

Construction and Characterization of pdgB Mutant—For construction of the pdgB mutant, genes Cj1121c to Cj1126c were PCR-amplified from C. jejuni NCTC 11168 using the primers: Cj1121cF (5'-CCTTACTATT-GGCATTAGAATAAGC-3') and Cj1126cR (5'-AAAAAAGCTTTGTTTAGGTCG-3'). The PCR product was polished with Fnu polymerase and then ligated into pPCR-Script Amp (Stratagene) according to the manufacturer's instructions. The ligation mixture was electroporated into electrocompetent E. coli DH10B and selected for on LB 5-gal agar (Sigma-Aldrich) with ampicillin. A blunt-ended kanamycin resistance cassette from pILL600 (29) was inserted into the filled-in XbaI restriction site of pdgB, generating pEAp26. The orientation of the cassette was confirmed to be non-polar by sequencing with the ckanB primer (5'-CTCGGGTTTCTAAGCATTAG-3'). DNA was sequenced using terminator chemistry and AmpliTaq cycle sequencing kits (Applied Biosystems) and analyzed on an Applied Biosystems 373A DNA sequencer. The mutated plasmid DNA was used for electroporation into C. jejuni CECT 11168 (30), and the kanamycin-resistant transformants were characterized by PCR to confirm that the incoming plasmid DNA had integrated by a double cross-over event.

Proteins were extracted from C. jejuni whole cells using 0.2 M glycine at pH 2.2 (17) and dialyzed against water. Samples were analyzed by two-dimensional PAGE using 11-cm pH 3–10 ReadyStrips (Bio-Rad Laboratories) as above and pre-cast 12 x 8-cm 8–16% gradient Criterion slab gels (Bio-Rad Laboratories). Gels were stained with colloidal Coomassie Blue, photographed, and then partially destained by washing in water. Proteins were transferred for 1 h at 207 mA onto PVDF membranes using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad). After blocking overnight, membranes were probed with a 1:500 dilution of H5:2 serotying serum followed by a 1:5000 dilution of goat anti-rabbit antiserum (Sigma-Aldrich) and developed with nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate (Roche Molecular Biochemicals).

**RESULTS**

**Purification and Characterization of PEB3—**PBE3 protein (Cj0289c) was identified in two-dimensional gels of a glycan extract by peptide mass fingerprinting, as a component of a group of spots focusing within a range of pH 9–10 (results not shown). PBE3 was purified from the extract by cation exchange chromatography, and re-fractionated on the same column, using a shallower NaCl gradient, resulting in the PEB3 appearing in three fractions (Fig. 1). SDS-PAGE analysis showed two bands, whose N-terminal sequences were determined following their transfer to a PVDF membrane. Ten cycles of sequencing identified the lower mass species as PEB3, whereas the higher mass, more abundant component, was also PEB3 with a minor sequence corresponding to PEB4 (Cj0596).

The mass spectrum and the reconstructed molecular mass profile for fraction #31 are presented in Fig. 2 (a and b). Three peaks were observed in the reconstructed mass profile. The peaks at 25,453 and 28,376 Da correspond well with the expected molecular masses of PEB3 (25,453 Da, Cj0289c) and PEB4 (28,377 Da, Cj0596) respectively, without signal peptides. To identify the protein of mass 28,861 Da, CapLC-MS/MS analysis was carried out on the tryptic digest of this fraction. All but one of the peptides identified could be assigned to PEB3 or PEB4, in accord with the N-terminal sequence data. MS/MS analysis of the unidentified ion (Fig. 3a) clearly identifies it as a glycopeptide. A fragmentation series composed of sequential losses of HexNAc (203 Da) and a single Hex (162 Da) can be observed in this spectrum. The tryptic peptide was identified as 62DFNVS27 from PEB3. The residue mass of the oligosaccharide portion of this glycopeptide is 1406 Da, which corresponds well with the difference in the molecular weights of PEB3 and
Characterization of the Glycopeptide Linkage—The PEB3 tryptic peptide to which the oligosaccharide is attached contains sites for both N- and O-linkage, i.e. Asn and Ser. Therefore, it was necessary to carry out further experiments to determine the nature of this linkage. Previously, we have used β-elimination to remove O-linked carbohydrates from the flagellin of C. jejuni 81–176 (11). However, this procedure failed to remove the oligosaccharide in this instance. This was our first indication that the oligosaccharide is N-linked. This was confirmed by MS/MS analysis of the singly protonated fragment ion at m/z 307.0 produced by front-end collision-induced dissociation of the intact glycopeptide (Fig. 3b). This ion is composed of the tryptic peptide plus the unusual 228-Da sugar only. An ion was observed at m/z 605.1, which could only be assigned to the b9 fragment ion plus the 228-Da sugar moiety. No fragment ions were observed, suggesting that the carbohydrate is linked to the serine residue. All of this evidence strongly suggests that the oligosaccharide is linked to the peptide at Asn70. Interestingly, this peptide contains the eukaryotic N-linkage consensus sequon, Asn-Xaa-Ser.
or bind nonspecifically to the column. This conclusion is supported by the failure of these protein spots to react with the SBA/alkaline phosphatase conjugate following two-dimensional PAGE and electroblotting (Table I).

### Preparation of Glycopeptides

The mixed glycoprotein sample was subjected to two rounds of Pronase digestion, and the products were separated by gel filtration on BioGel P4 (Fig. 5a). The carbohydrate-containing fractions were located by mass spectrometry and, after NMR studies (see below), the sample was re-purified on BioGel P2 (Fig. 5b). The final purified sample was composed mainly of the oligosaccharide linked to a single Asn (Fig. 5c). There was no evidence of any variation in the glycan component. The yield of glycopeptides was estimated at around 200 μg.

### NMR Spectroscopy

With the use of selective methods, it was possible to work with the impure P4 sample. The complete structure of the N-Linked Glycan of C. jejuni 42534 was elucidated.

---

### Table I

Identification of proteins from two-dimensional gels

| Cj gene | Annotation | Number of sequons | SBA staining | pgIB mutant |
|---------|------------|------------------|--------------|-------------|
| Cj0114e | Probable periplasmic protein | 2S, 3T | + | + |
| Cj0143c | Periplasmic solute binding protein for ABC transport system | 1S, 1T | + | + |
| Cj0147c | Thioredoxin (TrxA) | – | – | – |
| Cj0169c | Superoxide dismutase (Fe; SodB) | – | – | – |
| Cj0175e | Putative iron uptake ABC transport system periplasmic iron binding protein | 1S, 2T | + | + |
| Cj0200c | Probable periplasmic protein | 1T | + | + |
| Cj0238c | Probable integral membrane protein | 5S, 1T | – | – |
| Cj0269c | Major antigenic peptide (PEE3) | 2S | + | + |
| Cj0332c | Nucleoside diphosphate kinase (NdK) | – | – | – |
| Cj0334c | Alkyl hydroperoxide reductase (AhpC) | – | – | – |
| Cj0376c | Probable periplasmic protein | 1S, 1T | + | + |
| Cj0415c | Putative oxidoreductase sub-unit | 3T | + | + |
| Cj0420c | Probable periplasmic protein | 1T | ± | + |
| Cj0485c | Translation elongation factor EF-G (FusA) | 1T | ± | + |
| Cj0511c | Probable secreted proteinase | 2S, 2T | + | + |
| Cj0638c | Inorganic pyrophosphatase (Ppa) | – | – | – |
| Cj0694c | Probable periplasmic protein | 4S, 2T | + | + |
| Cj0715c | Transthrethin-like periplasmic protein | – | – | – |
| Cj0779c | Thioredoxin peroxidase (Tpx) | 1S | + | + |
| Cj0835c | Aconitate hydratase (AconB) | 1S, 2T | + | + |
| Cj0843c | Putative secreted transglycosylase | 5S, 3T | + | + |
| Cj0906c | Probable periplasmic protein | 2S, 2T | + | + |
| Cj0944c | Probable periplasmic protein | 1S, 1T | + | + |
| Cj0998c | Probable periplasmic protein | 1S, 1T | + | + |
| Cj1018c | Branched-chain amino-acid ABC transport system periplasmic binding protein | 1S, 2T | + | + |
| Cj1032c | Probable membrane fusion component of efflux system | 2T | + | + |
| Cj1118c | Translation elongation factor EF-Ts (Tsf) | – | – | – |
| Cj1214c | Hypothetical protein | 1S | ± | ± |
| Cj1221c | 60-kDa chaperonin (Cpn60; GroEL) | 1S, 1T | + | + |
| Cj1245c | Probable periplasmic protein | 5S, 2T | + | + |
| Cj1340c | Probable periplasmic protein | 2T | – | – |
| Cj1444c | Putative capsule polysaccharide export system periplasmic protein (KpsD) | 3S, 2T | + | + |
| Cj1496c | Probable periplasmic protein | 1S, 1T | + | + |
| Cj1534c | Probable bacterioferritin | – | – | – |
| Cj1565c | Paralyzed flagellum protein (PfaA) | 5S, 2T | + | + |
| Cj1643c | Putative periplasmic protein | 3S, 1T | + | + |
| Cj1659c | Periplasmic protein (P19) | – | – | – |
| Cj1670e | Probable periplasmic protein (CgpA) | 4S, 2T | + | + |

* S, Asn-Xaa-Ser sequons; T, Asn-Xaa-Thr sequons.
* Reactivity with SBA in Western blots of two-dimensional gels.
* Proteins with changed spot position and/or immunoreactivity in two-dimensional gels, in the pgIB mutant. Cj1018c and Cj1643 were not isolated by SBA chromatography.
* Glycopeptides observed by CapLC-MS/MS.
* Identified from one-dimensional gel.
resonance assignment of the sugar moiety was done with this sample for fear of loss of glycopeptide upon further purification. After a second purification using a P2 column, 25% of the glycopeptide was lost as judged by measurement of the signal-to-noise ratio for the P4- and P2-purified samples. An HMQC analysis of pooled material from P4 on BioGel P2 fine grade. In both fractionations, glycoprotein-containing fractions were identified by MS and pooled as indicated by the bars. c, ESI-MS spectrum of fraction 10 from B above. The doubly protonated ion (MH$^+$) at m/z 770.5 corresponds to the heptasaccharide linked to Asn. A number of larger ions are also observed and are due to the addition of a second amino acid residue. The amino acid compositions of the major glycan-containing ions are indicated on the spectrum.

FIG. 5. Purification of glycopeptides from a Pronase digest of the SBA affinity chromatography product. a, size exclusion chromatography on BioGel P4 200 mesh of the Pronase digest. b, re-fractionation of pooled material from P4 on BioGel P2 fine grade. In both fractionations, glycoprotein-containing fractions were identified by MS and pooled as indicated by the bars. c, ESI-MS spectrum of fraction 10 from B above. The doubly protonated ion (MH$^+$) at m/z 770.5 corresponds to the heptasaccharide linked to Asn. A number of larger ions are also observed and are due to the addition of a second amino acid residue. The amino acid compositions of the major glycan-containing ions are indicated on the spectrum.

A value of $J_{1^H,2^H}$ of 3.6 ± 0.2 Hz, the strong H1–H2 NOE (Fig. 7), and $J_{1^H,1^H}$ coupling pattern, which included a small coupling to H4 (Fig. 7), showed that these units had the $\alpha$-D-galactopyranosyl configuration. Although the $e1$ and $f1$ anomeric resonances could be selectively excited with a narrow bandwidth of 10 Hz, the spectra for simultaneous excitation of $e1$ and $f1$ are shown in Fig. 7b. Resonances $f2$ to $f4$ were also detected for the one-dimensional TOCSY of $g1f2$ (Fig. 7a). A chemical shift for C2 near 51 ppm was indicative of an acetamido group. The (C2, H4) HMBC correlations were used to
assign the C2 resonances of the five GalNAc residues. The one-dimensional TOCSY-NOESY(H1, H4) was used to detect the NOE between H4 and H5 and thus assign the H5 resonances (23). The one-dimensional TOCSY experiments on the H5 resonances were then used to detect the H6a resonances (not shown). Integration of the P2-purified sample (Fig. 6b) indicated seven NAc groups, five of those corresponding to the five GalNAc residues. Comparison of the $^{13}$C chemical shifts of the GalNAc units with those of $\beta$-b-GalNAc indicated that residues $a$, $c$, $e$, and $f$ were linked at O-4 due to downfield shifts for C4 (32). Residue $f$ had a branch point at O-3 as established by a downfield shift for C3. Residue $d$ was terminal due to similar $^{13}$C chemical shifts for C2 to C6 with those of $\alpha$-b-GalNAc (32).

Residue $b$ was assigned to $\beta$-b-bacillosamine (2,4-diacetamido-2,4,6-trideoxy-$\beta$-b-glucopyranose). For residues $b$ and $c$, the anomeric resonances partially overlapped. For the one-dimensional TOCSY of the $b1$ and $c1$ resonances with a mixing time of 144 ms (Fig. 7e), broad resonances could be identified up to a CH$_3$ resonance (b6) at 1.14 ppm. For the one-dimensional TOCSY for b6 at 1.14 ppm, with a mixing time of 66 ms (Fig. 7d), peaks up to b4 were observed. A series of one-dimensional TOCSY with different mixing time from 30 ms to 144 ms were done to assign the peaks. For residue b, the broad peaks do not provide the coupling constants. However, for the one-dimensional TOCSY of b1 with a mixing time of 144 ms, resonances up to H6 could be observed, similar to the one-dimensional TOCSY for H1 of $\beta$-Glc (residue g). Hence, such efficient transfer was indicative of large coupling constants typical of a $\beta$-glucopyranosyl configuration. The b1–b3 and b1–b5 NOE values observed in Fig. 7k were also typical of the $\beta$ anomeric configuration. Chemical shifts for C2 and C4 at 55 and 58 ppm were indicative of acetamido groups, in accord with the presence of seven NAc groups in the structure (Fig. 6a). The chemical shift for C1 at 79 ppm and H1 at 5.1 ppm indicated an N-linked anomeric similar to that found for $\beta$-GlcNAc-Asn (33). Comparison of the chemical shifts of residue b with those of 2,4-diamino-2,4,6-trideoxy-$\beta$-b-glucopyranose found in other structures indicated that Bac was linked at O-3, the only possible glycosidation site (34–36).

The absolute configuration of residue b was obtained by NOEs as previously described for another structure containing bacillosamine (34). A strong NOE observed between the CH$_3$ resonance at 1.14 ppm (b6) and the NAc at 1.95 ppm was due to the close proximity of the CH$_3$ group to the NAc-CH$_3$ group at C4. Because the NAc resonance at 1.95 ppm was isolated from the other NAc resonances, it could be selectively excited. A strong NOE was observed between this NAc resonance and the $\alpha$-b-GalNAc H1 resonance of residue $a$, due to the a(1→3)b linkage as shown below. This NOE can only occur if residue $b$ has the b configuration, because the NAc group at C4 of residue $b$ is in close proximity (3 Å) to the anomeric proton of residue $a$. This NOE is not possible if residue $b$ has the l configuration, because the H1a/b-NAc interproton distances are greater than 5 Å.

The sequence was determined by one-dimensional NOESY experiments (Fig. 7). The strong c1→f4 NOE and smaller c1→f6 and c1→f8 NOEs established the c(1→4)f linkage. The strong d1→c4 NOE and a smaller d1→c6 indicated the d(1→4)e sequence. The e1→a4, f1→e4, and g1→f3 established the e(1→4)k, f(1→4)e, and g(1→3)f linkages. The a1→b3 NOE (Fig. 7i) established the $a$→b sequence. The structure is shown in Fig. 6a. The sequence is in accord with the glycosidation shift observed for the linkage carbons (see above).

**Characterization of a pgIB Mutant—**A pgIB mutant was constructed by cassette mutagenesis (Fig. 8). Glycine extracts of the mutant cells demonstrated dramatic changes in protein immunoreactivity by two-dimensional PAGE with Western blotting with HS:2 serotyping sera (Fig. 8, b and e). Several of the proteins that showed a change in the mobility and/or immunoreactivity on two-dimensional gels were identified by mass fingerprint analysis (Table I). The protein identifications are in agreement with those identified by SBA lectin affinity chromatography, providing further evidence that the proteins reactive with the GalNAc lectin are glycosylated by the Pgl pathway. The full set of SBA-reactive proteins was not observed in this experiment, because it was performed on whole glycine extracts and at different pH range, hence Figs. 4 and 8 are not directly comparable. When PEB3 was purified from the pgIB mutant by ion-exchange chromatography as described above, analysis by mass spectrometry showed that the protein completely lacked the glycans (Fig. 2, c and d).

To show that the pgIB mutation only affected the glycosylation of the glycoproteins, analyses were performed of the lipoo-

---

**Table II**

Chemical shifts (ppm) of the C. jejuni Asn-linked glycopeptide

| Residue | C1 | C2 | C3 | C4 | C5 | C6 | H6 |
|---------|----|----|----|----|----|----|----|
| (a) α-GalNAc | 98.0 | 50.7 | 68.0 | 77.6 | 72.7 | 62.2 | 3.90 |
| (b) β-Bac | 79.0 | 54.5 | 76.3 | 58.1 | 75.0 | 17.6 | |
| (c) α-GalNAc | 98.3 | 51.6 | 67.9 | 77.5 | 71.8 | 60.9 | 3.62 |
| (d) α-GalNAc | 99.7 | 51.4 | 68.4 | 69.6 | 72.3 | 60.7 | |
| (e) α-GalNAc | 99.6 | 51.6 | 67.9 | 77.5 | 72.4 | 60.7 | 3.61 |
| (f) α-GalNAc | 99.9 | 50.7 | 77.5 | 75.8 | 72.5 | 60.5 | 3.53 |
| (g) β-Glc | 106.1 | 74.1 | 76.9 | 71.1 | 77.2 | 62.1 | 3.68 |
ligosaccharide and capsular polysaccharide of the mutant. Deoxycholate-PAGE of proteinase K digests and mass spectrometry showed that the mass of the mutant LOS core was identical to that of the wild-type (results not shown). In addition, identical capsular repeats were visible in the extracts of the cells of the wild-type and the isogenic mutant on deoxycholate-PAGE. To further demonstrate that the capsule was unaltered, we examined the polysaccharide by HR-MAS NMR, which has been developed to examine capsular polysaccharide ligosaccharide and capsular polysaccharide of the mutant. The mass data reported for the PEB3 analysis showed the substituted Asn was followed by Val-Ser, suggesting the sequon for glycosylation was the same as in eukaryotes, Asn-Xaa-Ser/Thr.

In all, 38 proteins were identified in one-dimensional and two-dimensional PAGE of the SBA affinity chromatography product, by MS of the products of in-gel tryptic digestions (Table I). Of these, 22 were found to react with SBA or immune sera in membrane blots. The remainder generally gave weak spots upon silver staining but still failed to react when blots of gels from higher protein loads were re-tested. Consistent with this behavior, all but one of the sequences of these proteins completely lacked any Asn-Xaa-Ser/Thr sequons. Although they might therefore be considered nonspecifically absorbed proteins, the elution agent, 0.1 M GalNAc, is a mild one. Hence, it is possible that some of them form complexes with one or other of the glycoproteins. The glycoproteins are predominantly annotated as periplasmic proteins, which suggests the glycosylation machinery is specific for this compartment. The majority of the proteins gave rise to several spots on the two-dimensional PAGE indicating multiple glycoforms, which in the case of the purified PEB3 includes the non-glycosylated form. In addition, several other periplasmic proteins observed in the proteome map of C. jejuni have sequons but did not appear in the SBA extract. Hence, the glycosylation is generally incomplete. The PEB3 data showed Asn-Xaa-Ser to be the probable sequon. However, MS/MS analysis of the gel-isolated glycoproteins digestates indicates that the alternative N-linkage sequon, Asn-Xaa-Thr, is being utilized in at least three other glycoproteins (Cj0114, Cj0209c, and Cj1670c). In addition, two other glycoproteins identified in the lectin-affinity isolate, Cj0415 and Cj1032, have Asn-Xaa-Thr sequons but no Asn-Xaa-Ser ones. The Thr sequon occurs relatively often in the C. jejuni glycoprotein sequences, at 43% of the putative Asn sequons. That the glycan structure is an immunodominant feature was evident from blots of two-dimensional gels of the SBA-purified glycoproteins with immune sera, which were virtually identical to blots with SBA, and from the loss of protein spots in the blots of the pglB mutant (Fig. 8).

It has previously been pointed out (3, 16) that the C. jejuni genome contains a gene, pglB, which shows homology to the gene encoding the STT3 subunit of the S. cerevisiae oligosaccharyltransferase (P39007; 14% identity and 26% similarity). This enzyme is responsible for the block transfer of glycans to Asn residues to form glycoproteins. Furthermore, there is an unannotated gene in a member of the Archaea, Methanobacterium thermoautotrophicum (3), that is more homologous to either of the above than they are to each other (AE000942; 15% identity and 27% similarity to STT3; 17% identity and 26% similarity to pglB). Hence, this pathway appears to be present in all three domains. The establishment here of the presence of N-linked glycans in C. jejuni suggests that PglB may also function as an oligosaccharyltransferase, and this was further confirmed by the complete loss of glycan from PEB3 in the pglB mutant. Although the genetic organization of the locus suggests many of the pgl genes are co-transcribed (3), our results suggest that the phenotype of the mutant does not reflect polar effects on downstream genes. However, PglB differs from STT3 in its putative specificity being for transfer to Asn of a heptasaccharide with a Bac-reducing end group, instead of Glc3Man9GlcNAc2, although both oligosaccharides may be transferred from lipid carriers. Although PglB has not been shown to form a complex with other proteins in the manner of STT3, the two proteins have similar membrane topologies and homologous C-terminal domains (16). From the point of view of N-glycosylation, therefore, the periplasmic space would appear

DISCUSSION

By means of mass spectrometry and HR-MAS nano-NMR techniques (11, 37), the structure of the Asn-linked glycan was established as a heptasaccharide comprising a linear chain of one β-Bac and five β-GalNAc residues with a β-Glc branch (Fig. 6a). The non-reducing terminal GalNAc residue would be responsible for the reactivity of C. jejuni glycoproteins with the lectin, SBA (15). Bacillosamine (Bac), a diaminotrideoxyhexosamine, has been found in many prokaryotes (36), most recently in Pseudomonas reactans (35). An N-linkage between Bac and Asn residues has not been reported before, although it is possible that an O-linkage of Bac to Ser is present in the pilin of N. meningitidis (5, 6). The mass data reported for the O-linked sugar are consistent with Bac (38), but its stereochemistry has not been reported. No evidence was found for any variation in the C. jejuni glycopeptide structure within the glycan moiety, although, as was to be expected, there were glycopeptides with various additional amino acids in the Pronase product (Fig. 5c). The PEB3 analysis showed the substituted Asn was followed by Val-Ser, suggesting the sequon for glycosylation was the same as in eukaryotes, Asn-Xaa-Ser/Thr.

In all, 38 proteins were identified in one-dimensional and two-dimensional PAGE of the SBA affinity chromatography product, by MS of the products of in-gel tryptic digestions (Table I). Of these, 22 were found to react with SBA or immune sera in membrane blots. The remainder generally gave weak spots upon silver staining but still failed to react when blots of gels from higher protein loads were re-tested. Consistent with this behavior, all but one of the sequences of these proteins completely lacked any Asn-Xaa-Ser/Thr sequons. Although they might therefore be considered nonspecifically absorbed proteins, the elution agent, 0.1 M GalNAc, is a mild one. Hence, it is possible that some of them form complexes with one or other of the glycoproteins. The glycoproteins are predominantly annotated as periplasmic proteins, which suggests the glycosylation machinery is specific for this compartment. The majority of the proteins gave rise to several spots on the two-dimensional PAGE indicating multiple glycoforms, which in the case of the purified PEB3 includes the non-glycosylated form. In addition, several other periplasmic proteins observed in the proteome map of C. jejuni have sequons but did not appear in the SBA extract. Hence, the glycosylation is generally incomplete. The PEB3 data showed Asn-Xaa-Ser to be the probable sequon. However, MS/MS analysis of the gel-isolated glycoproteins digestates indicates that the alternative N-linkage sequon, Asn-Xaa-Thr, is being utilized in at least three other glycoproteins (Cj0114, Cj0209c, and Cj1670c). In addition, two other glycoproteins identified in the lectin-affinity isolate, Cj0415 and Cj1032, have Asn-Xaa-Thr sequons but no Asn-Xaa-Ser ones. The Thr sequon occurs relatively often in the C. jejuni glycoprotein sequences, at 43% of the putative Asn sequons. That the glycan structure is an immunodominant feature was evident from blots of two-dimensional gels of the SBA-purified glycoproteins with immune sera, which were virtually identical to blots with SBA, and from the loss of protein spots in the blots of the pglB mutant (Fig. 8).

It has previously been pointed out (3, 16) that the C. jejuni genome contains a gene, pglB, which shows homology to the gene encoding the STT3 subunit of the S. cerevisiae oligosaccharyltransferase (P39007; 14% identity and 26% similarity). This enzyme is responsible for the block transfer of glycans to Asn residues to form glycoproteins. Furthermore, there is an unannotated gene in a member of the Archaea, Methanobacterium thermoautotrophicum (3), that is more homologous to either of the above than they are to each other (AE000942; 15% identity and 27% similarity to STT3; 17% identity and 26% similarity to pglB). Hence, this pathway appears to be present in all three domains. The establishment here of the presence of N-linked glycans in C. jejuni suggests that PglB may also function as an oligosaccharyltransferase, and this was further confirmed by the complete loss of glycan from PEB3 in the pglB mutant. Although the genetic organization of the locus suggests many of the pgl genes are co-transcribed (3), our results suggest that the phenotype of the mutant does not reflect polar effects on downstream genes. However, PglB differs from STT3 in its putative specificity being for transfer to Asn of a heptasaccharide with a Bac-reducing end group, instead of Glc3Man9GlcNAc2, although both oligosaccharides may be transferred from lipid carriers. Although PglB has not been shown to form a complex with other proteins in the manner of STT3, the two proteins have similar membrane topologies and homologous C-terminal domains (16). From the point of view of N-glycosylation, therefore, the periplasmic space would appear

---

2 St. Michael, F., Szymanski, C. M., Li, J., Chan, K. H., Khieu, N. H., Larocque, S., Wakarchuk, W. W., Brisson, J.-R., and Monteiro, M. A. (2002) Eur. J. Biochem. 269, in press.
to be functionally similar to the endoplasmic reticulum.

In our hands, the structures of the LOS and capsule were unaffected by pglB mutation, as previously found (3, 15). There are effects of pgl gene mutation on the immunological properties of flagellin (3), but these do not appear to be attributable to loss of the glycan from it. While the two structural flagellin proteins have seven or eight potential glycosylation sites, respectively, only O-linked pseudaminic acid variants have been found on these proteins so far (11). The changes appear therefore to be due to some indirect effect, and it is noteworthy that the list of glycoproteins (Table I) includes one annotated as being involved in flagellin biosynthesis, Cj1565c, pflA the paralogue factor (39).

The pgl locus of C. jejuni is very similar to one in N. meningitidis that is responsible for the O-glycosylation of pilin (5, 6), as shown in Fig. 8. Although in general the annotation of carbohydrate-related genes cannot be very precise, nevertheless there are some interesting connections between the annotations of genes in the loci and the glycan structures. The pilin of N. meningitidis is O-substituted at Ser87 with Gal-1,3-Gal-1,3-di-oligo-D-glucosamindhexose (38). The conformation of the latter sugar has not been reported, but it could well be bacillosamine. In agreement with the amino acid linkages in the two structures, there is no homolog of Cj pglB in N. meningitidis, but three genes that are plausible candidates for Bac synthesis genes are present in both loci, Cj pglF (Nm pglD) a dehydratase, Cj pglE (Nm pglC) an aminotransferase, and Cj pglD (Nm pglB) an acetyltransferase (annotations from Parkhill et al. (13)). Cj pglC is a homolog of the second section of the apparently fused gene Nm pglB, which has been suggested to be a glycosyltransferase responsible for putting the first sugar, i.e. bacillosamine, onto a lipid carrier (6). This infers that both N- and O-linked structures are pre-assembled on lipids before attachment to protein, rather than the latter being assembled processively in the manner of eukaryotic O-linked glycoproteins (40).

This analysis suggests that much of the homology between the two loci is due to Bac-related genes. It is therefore surprising that N. gonorrhoeae also contains a set of these genes (41), when the reported glycan substitution on its pilin is Gal-1,3-GlcNAc (42). The glycosyltransferases that assemble the rest of the structures are less easy to assign. The Gal-1,3 linkage in the N. meningitidis structure has been shown to be formed by the Nm pglA product, which lies well outside the locus (5). This is homologous to the Cj pglA, which is therefore a candidate for the transferase forming the GalNAc-1,3-Bac linkage. Nm pglA is also highly homologous to an N. gonorrhoeae gene pglA, which could therefore be responsible for forming the Gal-1,3-GlcNAc linkage.

The other putative glycosyltransferase genes in C. jejuni are Cj1127–1129 or pglH, pglI, and pglJ. These have relatively general annotations that do not permit assignment to GalNAc-1,4- or GlcB1,3-transferase functions. The lack of homology among them and being only these three candidate genes in the locus for the remaining five glycoside linkages argue against postulating four independent GalNAc-1,4 transferases for assembly of the remainder of the structure. One such transferase could assemble the GalNAc chain with the oligosaccharyltransferase being sensitive to the size of the assembled glycan. Finally, there is a nearby gene, Cj1130c, an ABC transporter wlaB, which may be involved in transfer of the lipid-attached glycan across the inner membrane.

The bacterial glycoproteins that have been identified previously are predominantly single types of protein in each bacterium, such as flagellin or S-layers (1, 2), whose biological function is therefore known. In C. jejuni, we have demonstrated that multiple glycoproteins contain the same glycan, suggesting that several biological functions could potentially be affected. In addition, the glycoproteins, such as PE3, can exist in both modified and unmodified forms. The ability to vary protein glycosylation has been described for M. tuberculosis Apa glycoproteins, which show differences in delayed-type hypersensitivity reactions and T-lymphocyte stimulation related to the extent of protein glycosylation (43, 44). Other biological roles that carbohydrate modifications on bacterial glycoproteins have been shown to affect include adhesion (45), protection against proteolytic cleavage (46), solubility (47), antigenic variation (48), and protective immunity (49). In C. jejuni, alteration of the N-linked glycosylation pathway by mutation in the pgl locus has already been shown to influence adherence, invasion, colonization, and immunogenicity (3, 4). We are currently investigating the biological role(s) of the glycan modifications in C. jejuni pathogenesis.

Acknowledgments—We thank Dr. Warren W. Wakarchuk and Dr. Michel Gilbert for their valuable comments on the manuscript, Dr. Jianjian Li for the mass spectrometry of LOS, Dr. John Nash for primer design, and Dr. Kris Rahn, Health Canada, Guelph, Ontario, for HS:2 antisera.
34. Hermansson, K., Perry, M. B., Altman, E., Brisson, J. R., and Garcia, M. M. (1993) *Eur. J. Biochem.* **212**, 801–809
35. Melinare, A., Evidente, A., Sante, I. N., Lanzaeta, R., Lo, C. P., Mancino, A., and Parrilli, M. (2002) *Carbohydr. Res.* **337**, 467–471
36. Schäffer, C., Scherf, T., Christian, R., Kosma, P., Zayni, S., Messner, P., and Sharon, N. (2001) *Eur. J. Biochem.* **268**, 857–864
37. Gallego, R. G., Blanco, J. L., Thijssen-van Zuylen, C. W., Gotfredsen, C. H., Voshol, H., Duus, J. O., Schachner, M., and Vliegenthart, J. F. (2001) *J. Biol. Chem.* **276**, 30834–30844
38. Stimson, E., Pirjol, M., Makepeace, K., Dell, A., Morris, H. R., Payne, G., Saunders, J. R., Jennings, M. P., Barker, S., and Panico, M. (1995) *Mol. Microbiol.* **17**, 1201–1214
39. Yao, R., Burr, D. H., Doig, P., Trust, T. J., Niu, H., and Guerry, P. (1994) *Mol. Microbiol.* **14**, 883–893
40. Gemmill, T. R., and Trimble, R. B. (1999) *Biochim. Biophys. Acta* **1426**, 227–237
41. Kahler, C. M., Martin, L. E., Tzeng, Y. L., Miller, Y. K., Sharkey, K., Stephens, D. S., and Davies, J. K. (2001) *Infect. Immun.* **69**, 3597–3604
42. Parge, H. E., Forrest, K. T., Hickey, M. J., Christensen, D. A., Getzoff, E. D., and Tainer, J. A. (1995) *Nature* **378**, 32–38
43. Romain, F., Horn, C., Pescher, P., Namane, A., Riviere, M., Puzo, G., Barzu, O., and Marchal, G. (1999) *Infect. Immun.* **67**, 5567–5572
44. Horn, C., Namane, A., Pescher, P., Riviere, M., Romain, F., Puzo, G., Barzu, O., and Marchal, G. (1999) *J. Biol. Chem.* **274**, 32023–32030
45. Benz, I., and Schmidt, M. A. (2001) *Mol. Microbiol.* **40**, 1403–1413
46. Herrmann, J. L., Delahay, E., Gallagher, A., Robertson, B., and Young, D. (2000) *FEBS Lett.* **473**, 358–362
47. Marceau, M., and Nassif, X. (1999) *J. Bacteriol.* **181**, 656–661
48. Doig, P., Kirseh, N., Guerry, P., and Trust, T. J. (1996) *Mol. Microbiol.* **19**, 379–387
49. Guerry, P., Doig, P., Kirseh, N., Burr, D. H., Kirseh, N., and Trust, T. J. (1996) *Mol. Microbiol.* **19**, 369–378
Structure of the N-Linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, *Campylobacter jejuni*

N. Martin Young, Jean-Robert Brisson, John Kelly, David C. Watson, Luc Tessier, Patricia H. Lanthier, Harold C. Jarrell, Nicolas Cadotte, Frank St. Michael, Erika Aberg and Christine M. Szymanski

*J. Biol. Chem.* 2002, 277:42530-42539. doi: 10.1074/jbc.M206114200 originally published online August 16, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M206114200

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 48 references, 13 of which can be accessed free at http://www.jbc.org/content/277/45/42530.full.html#ref-list-1