# The Genetic Bases for the Variation in the Lipo-oligosaccharide of the Mucosal Pathogen, *Campylobacter jejuni*

## BIOSYNTHESIS OF SIALYLATED GANGLIOSIDE MIMICS IN THE CORE OLIGOSACCHARIDE*

### Michel Gilbert‡, Marie-France Karwaski, Stéphane Bernatchez, N. Martin Young, Eduardo Taboada, Joseph Michniewicz, Anna-Maria Cunningham, and Warren W. Wakarchuk

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

We have compared the lipo-oligosaccharide (LOS) biosynthesis loci from 11 *Campylobacter jejuni* strains expressing a total of 8 different ganglioside mimics in their LOS outer cores. Based on the organization of the genes, the 11 corresponding loci could be classified into three classes, with one of them being clearly an intermediate evolutionary step between the other two. Comparative genomics and expression of specific glycosyltransferases combined with *in vitro* activity assays allowed us to identify at least five distinct mechanisms that allow *C. jejuni* to vary the structure of the LOS outer core as follows: 1) different gene complements; 2) phase variation because of homopolymeric tracts; 3) gene inactivation by the deletion or insertion of a single base (without phase variation); 4) single mutation leading to the inactivation of a glycosyltransferase; and 5) single or multiple mutations leading to "allelic" glycosyltransferases with different acceptor specificities. The differences in the LOS outer core structures expressed by the 11 *C. jejuni* strains examined can be explained by one or more of the five mechanisms described in this work.

Many pathogenic bacteria have variable cell-surface glycoconjugates such as capsules in *Streptococcus* spp. and *Neisseria meningitidis* (1), lipopolysaccharides in Gram-negative bacteria (2), and glycosylated surface-layer proteins (3). In mucosal pathogens, the variability of cell-surface polysaccharides has been shown to play a major role in virulence (4). This variation is caused by the diversity of monosaccharide components and the linkages between them, derivatization with noncarbohydrate moieties, and in some cases, by the length and sequence of the repeating units. The variation of these glycan structures can sometimes be correlated with a specific gene complement, but it is probable that other genetic mechanisms are also employed to create variable cell-surface glycoconjugates. The DNA sequencing of the relevant genetic loci from multiple strains of a pathogen can provide insights into the genetic origins of important strain variable traits such as cell-surface glycoconjugates.

The mucosal pathogen *Campylobacter jejuni* has been recognized as an important cause of acute gastroenteritis in humans (5) and has been shown to have variable cell-surface carbohydrates that are associated with virulence (6, 7). Epidemiologi-

cal studies have shown that *Campylobacter* infections are more common than *Salmonella* infections in developed countries, and they are also an important cause of diarrheal diseases in developing countries. *C. jejuni* is also considered the most frequent antecedent infection to the development of Guillain-Barré syndrome, a form of neuropathy that is the most common cause of generalized paralysis since the eradication of poliomyelitis in developed countries (8). The core oligosaccharides of low molecular weight lipo-oligosaccharides (LOS)[1] of many *C. jejuni* strains have been shown to exhibit molecular mimicry of the carbohydrate moieties of gangliosides (Fig. 1). Terminal oligosaccharides identical to those of GM1a, GM2, GM3, GD1a, GD1c, GD3, and GT1a gangliosides have all been found in various *C. jejuni* strains (see Table I for references). Molecular mimicry of host structures by the saccharide portion of LOS is considered to be a virulence factor of various mucosal pathogens, which may use this strategy to evade the immune response (9). The molecular mimicry between *C. jejuni* LOS outer core structures and gangliosides has also been suggested to act as a trigger for autoimmune mechanisms in the development of Guillain-Barré syndrome (10).

Aspinall *et al.* (11–14) and Nam Shin *et al.* (15) determined the LOS outer core structures of representative *C. jejuni* reference strains of the Penner serotyping system. The Penner serotyping system of *C. jejuni* is based on heat-stable antigens, and it was proposed that the specificity is due to LOS and/or lipopolysaccharide-type molecules (16, 17). However, recent biochemical and genetic studies suggest that capsular polysaccharides account for Penner serotype specificity (6, 18). Because the loci responsible for capsule and LOS biosynthesis are distant in the *C. jejuni* genome (19) and intraspecies gene transfers are known to be frequent in *C. jejuni* (20, 21), it is possible that strains having the same Penner type could express different LOS outer cores. Consequently, we decided to associate the published LOS outer core structures (Fig. 1) with the specific strain identification numbers (ATCC, NCTC, etc.) rather than with the Penner types, although the latter are also provided for convenient reference (Table I).

The identification of the genes involved in LOS synthesis and the study of their regulation are of considerable interest for a better understanding of the pathogenesis mechanisms used by these bacteria. The availability of the complete genome sequence of *C. jejuni* NCTC 11168 (19) has facilitated the iden-

[1] The abbreviations used are: LOS, lipo-oligosaccharides; CMP-Neu5Ac, cytidine monophosphate-*N*-acetylneuraminic acid; FCHASE, 6-(5-fluorescein-carboxamido)-hexanoic acid succimidyl ester; ORF, open reading frame; Hep, heptose; nt, nucleotide. The abbreviated designations of glycolipids are according to IUPAC-IUC nomenclature.

tification of loci involved in the biosynthesis of cell-surface carbohydrates including LOS (22, 23). The genome sequence was also used to clone the corresponding LOS biosynthesis locus in other *C. jejuni* strains (24, 25), which allowed the identification of genes involved in the transfer of Gal, GalNAc, and *N*-acetylneuraminic acid (Neu5Ac or sialic acid) to the LOS outer core.

Because cell-surface structures such as the LOS are recognized as antigens by the host, it is therefore not surprising that microorganisms will modulate these structures to increase the chances of evading the immune system. The *C. jejuni* strains used in this study were shown to express a total of 8 different sialylated LOS outer cores (Fig. 1 and see Table I for references). The LOS biosynthesis loci of *C. jejuni* OH4384 and *C. jejuni* NCTC 11168 were found to have common genes as well genes unique to each strain (24), which provide a basis for differences in LOS outer cores. However, mechanisms other than differences in gene complement are involved in generating a variety of LOS outer cores. In the strain *C. jejuni* OH4382, the gene involved in the transfer of the GalNAc residue of the LOS outer core was shown to be inactive (a missing A nucleotide causes a premature translation stop). This results in the expression of a truncated LOS outer core when compared with strain OH4384 (13, 24). Parkhill *et al.* (19) showed that short homopolymeric nucleotide runs of variable length are commonly found in genes involved in the biosynthesis of *C. jejuni* carbohydrates, which provides a form of on/off regulation of these genes. Linton *et al.* (22) studied in detail a gene encoding a $\beta$-1,3-galactosyltransferase that occurs with either an 8- or a 9-G nucleotide tract which results in the expression of either a GM1a or a GM2 ganglioside mimic in *C. jejuni* NCTC 11168. We reported previously that the *cst-II* gene occurs as a mono-functional $\alpha$-2,3-sialyltransferase in *C. jejuni* ATCC 43446 (O:19 serostrain) and as a bi-functional $\alpha$-2,3-/$\alpha$-2,8-sialyltransferase in *C. jejuni* OH4384 that results in the expression of either a GD1a or GT1a mimic, respectively (24).

In this work we describe the mechanisms used by *C. jejuni* to generate various sialylated outer core structures. In addition to reporting other examples of on/off expression of genes due to variable homopolymeric tracts, we use enzymatic assays to show that amino acid substitutions are responsible for the expression of glycosyltransferases with different substrate specificities, a "strategy" that further expands the ability of *C. jejuni* to express various LOS outer cores.

EXPERIMENTAL PROCEDURES

*Bacterial Strains*—The *C. jejuni* strains used in this study are listed in Table I. The Penner type strains were obtained from the American Type Culture Collection. *C. jejuni* OH4382, OH4384 and NCTC 11168 were obtained from the Laboratory Center for Disease Control (Health Canada, Winnipeg, Manitoba, Canada). *C. jejuni* strains were grown on Mueller-Hinton medium under microaerobic conditions. *Escherichia coli* AD202 (CGSG 7297) was used to express the different cloned glycosyltransferases and was grown using 2YT agar or broth. The recombinant *E. coli* strains were incubated at 25 °C for a total of 24 h, with induction with 1 mM isopropyl-1-thio-$\beta$-D-galactopyranoside after 6 h for *cgtA* constructs and with 0.3 mM isopropyl-1-thio-$\beta$-D-galactopyranoside after 4.5 h for *cst-II* constructs.

*Basic Recombinant DNA Method*—Genomic DNA isolation from the *C. jejuni* strains was performed using the DNeasy Tissue kit (Qiagen Inc., Valencia, CA). Plasmid DNA isolation, restriction enzyme digestions, purification of DNA fragments for cloning, ligations, and transformations were performed as recommended by the enzyme supplier or the manufacturer of the kit used for the particular procedure. Long PCRs (>2 kb) were performed using the Expand™ long template PCR system as described by the manufacturer (Roche Molecular Biochemicals). PCRs to amplify specific ORFs were performed using the *Pwo* DNA polymerase as described by the manufacturer (Roche Molecular Biochemicals). Restriction and DNA modification enzymes were purchased from MBI Fermentas Inc. (Hanover, MD). Site-directed mu-

tagenesis of cst-II was performed using a two stage PCR mutagenesis protocol. Two separate PCR reactions were performed to generate two overlapping gene fragments that both contained the mutation due to either the 5′ or the 3′ primers. The two PCR products were used with the cst-II 5′ and 3′ primers to amplify the full-length mutated version of *cst-II*.

*Sequencing of the LOS Biosynthesis Loci*—The DNA sequences of the LOS biosynthesis loci of *C. jejuni* NCTC 11168 (GenBank™ accession number AL139077) and OH4384 (GenBank™ accession number AF130984) were used to design primers to amplify the LOS biosynthesis loci of the other strains described in this work. The primers were designed to obtain overlapping PCR products of 2–5 kb that covered completely each of the LOS locus. The PCR products were sequenced by "primer walking," and new primers were synthesized to amplify and sequence the regions that diverge significantly from the NCTC 11168 and OH4384 sequences. DNA sequencing was performed using an Applied Biosystems (Montreal) model 373 automated DNA sequencer and the manufacturer's cycle sequencing kit.

*Assays*—Protein concentration was determined using the bicinchoninic acid protein assay kit (Pierce). FCHASE-labeled oligosaccharides were prepared as described previously (26). Extracts were made by sonication, and the enzymatic reactions were performed at 37 °C for 5 min to 2 h. The $\beta$-1,4-*N*-acetylgalactosaminyltransferase was assayed using 0.5 mM Neu5Ac$\alpha$-2,3-Gal$\beta$-1,4-Glc-FCHASE, 1 mM UDP-GalNAc, 50 mM Hepes, pH 7, and 10 mM MnCl$_2$. The $\alpha$-2,3-sialyltransferase was assayed using 0.5 mM Gal-$\beta$-1,4-Glc-FCHASE, 0.2 mM CMP-Neu5Ac, 50 mM Hepes, pH 7.5, and 10 mM MgCl$_2$. The $\alpha$-2,8-sialyltransferase was assayed using 0.5 mM Neu5Ac$\alpha$-2,3-Gal$\beta$-1,4-Glc-FCHASE, 0.2 mM CMP-Neu5Ac, 50 mM Hepes, pH 7.5, and 10 mM MgCl$_2$. The CMP-Neu5Ac synthetase was assayed using CTP, Neu5Ac, Gal-$\beta$1,4-GlcNAc-FCHASE, and a purified fusion of the *N. meningitidis* $\alpha$-2,3-sialyltransferase (MalE-NST)[2] in a coupled assay that measured the production of Neu5Ac$\alpha$-2,3-Gal-$\beta$1,4-GlcNAc-FCHASE. The reaction mix included 0.5 mM Gal-$\beta$-1,4-GlcNAc-FCHASE, 3 mM CTP, 3 mM Neu5Ac, 4 milliunits of $\alpha$-2,3-sialyltransferase (MalE-NST), 100 mM Tris, pH 7.5, 10 mM MgCl$_2$, and 0.2 mM dithiothreitol. All the reactions were stopped by the addition of acetonitrile (25% final concentration) and were diluted with H$_2$O to get 10–15 $\mu$M final concentration of the FCHASE-labeled compounds. The samples were analyzed by capillary electrophoresis performed using the separation and detection conditions as described previously (27). The peaks from the electropherograms were analyzed using manual peak integration with the P/ACE Station software.

RESULTS

*Organization of the LOS Biosynthesis Loci in the Various C. jejuni Strains*—We have compared the LOS biosynthesis loci of 11 *C. jejuni* strains (Table I) that include 7 previously unpublished loci and extend our previous limited comparison of 4 *C. jejuni* strains that included 3 closely related O:19 strains (OH4382, OH4384, and ATCC 43446, the O:19 serostrain) and the genome strain NCTC 11168 (24). The LOS outer core structures were published for 10 of the 11 strains included in this study (Fig. 1 and Table I). The general organization of the LOS biosynthesis genes allows us to group these *C. jejuni* strains into three classes "A," "B," and "C" (see Fig. 2). The LOS biosynthesis loci of the six class A strains have 13 ORFs, whereas the LOS biosynthesis loci of the two class B strains and of the three class C strains have 14 ORFs. One gene (*orf11*) is found only in classes A and B, whereas three genes are unique to class C (*orf14, orf15,* and *orf16*). Proposed functions for each ORF are described in Table II.

The 11.5-kb DNA sequences of the LOS loci from the six class A strains can be aligned with only minor gaps, the longest being 6 bp. The overall DNA sequence identity is 91% between the six A strains. However, the level of conservation observed in pairwise alignments varies considerably. As reported previously (24) the three O:19 strains (ATCC 43446, OH4382, and OH4384) are closely related. There is only one base difference (a missing A at position 71 of *orf5*) between the LOS locus of

---

TABLE I
*Description of the C. jejuni strains used in this work*

| Strain no. | Penner serotype | LOS outercore type[a] (ganglioside mimic) | Reference for the LOS outer core structure | LOS biosynthesis locus class (GenBank™ accession no.) |
|---|---|---|---|---|
| ATCC 43438 | O:10 (type strain) | I (GD1c) | 15 | A (AF400048) |
| ATCC 43432 | O:4 (type strain) | II (GD1a) | 11 | A (AF215659) |
| ATCC 43446 | O:19 (type strain) | II (GD1a) | 13 | A (AF167344) |
| OH4384 (GBS strain) | O:19 | III (GT1a) | 13 | A (AF130984) |
| OH4382 (GBS strain) | O:19 | IV (GD3) | 13 | A (AF167345) |
| ATCC 43460 | O:41 (type strain) | ? | None | A (AY044868) |
| ATCC 43449 | O:23 (type strain) | V (GM2) | 11 | B (AF401529) |
| ATCC 43456 | O:36 (type strain) | V (GM2) | 11 | B (AF401528) |
| NCTC 11168 ("Genome" strain) | O:2 | VI (GM1a) | 22[b] | C (AL139077) |
| ATCC 43429 | O:1 (type strain) | VII (GM2) | 11 | C (AY044156) |
| ATCC 43430 | O:2 (type strain) | VIII (GM3) | 12 | C (AF400047) |

[a] See Fig. 1 for the description of the structures.
[b] Linton *et al.* (22) and St. Michael, F., Gilbert, M., Syzmanski, C., Chan, K. H., Wakaschuk, W. W., and Monteiro, N., unpublished data.

OH4382 and OH4384. There are 68 base differences (20 amino acid differences) between ATCC 43446 (O:19 serostrain) and OH4384. The LOS locus from *C. jejuni* ATCC 43438 (O:10 serostrain) is primarily responsible for decreasing the overall degree of conservation among the A class strains. When the ATCC 43438 strain is excluded from the class A alignment, the overall DNA sequence identity increases to 96.5%. The highest level of divergence between the LOS locus of ATCC 43438 and the other class A strains is found between nt 4500 and 5700 (66% DNA sequence identity), a region that spans both the *orf5* and *orf6* which encode a β-1,4-*N*-acetylgalactosaminyltransferase and a β-1,3-galactosyltransferase, respectively.

The 12.4-kb LOS biosynthesis locus of the two class B strains (ATCC 43449, the O:23 serostrain, and ATCC 43456, the O:36 serostrain) shows 95.2% DNA sequence identity in a full-length pairwise alignment. However, the sequence identity is only 65.3% in the region from nt 4500 to 5700, whereas it is above 98% in the rest of the locus. It is noticeable that this region corresponds to the same region that was found to diverge considerably between ATCC 43438 and the other class A strains. In fact, ATCC 43438 and ATCC 43449 share 98% DNA sequence identity in the nt 4500–5700 region, whereas the other class A strains and ATCC 43456 share 99% DNA identity in that region.

Class B appears to be an evolutionary intermediate between classes A and C because it has two copies of *orf5*, with one of them (*orf5-I*) more similar to *orf5* from class A (96% DNA sequence identity) and the second copy (*orf5-II*) more similar to *orf5* from class C (85% DNA sequence identity). The *orf5-I* in the class B is inactive because of premature translational termination after 28 codons in ATCC 43449 and after 86 codons in ATCC 43456. Transcription reinitiation of *orf5-I* would theoretically be possible, but a similar frameshift mutation was described in *orf5* of OH4382 and resulted in the expression of a truncated LOS, consistent with the absence of active β-1,4-*N*-acetylgalactosaminyltransferase (24). The *orf5-II* is located just upstream of *orf10* in the class B (Fig. 2). Although *orf5-II* and *orf10* are separate ORFs in class B, they are found as an in-frame single ORF (*orf5/10*) in class C as reported previously (24). A genetic rearrangement is presumed to have occurred that led to the fusion of these two ORFs in class C.

The level of DNA sequence conservation among the loci from the three class C strains (ATCC 43429, the O:1 serostrain, ATCC 43430, the O:2 serostrain and NCTC 11168) is very high with a maximum of 18-base differences between them using pairwise comparisons across the whole 13.5-kb sequence. We describe below how some of the minor DNA sequence differences are responsible for the different LOS outer cores expressed by the three class C strains.

Comparisons among the three classes are more easily made

by aligning the corresponding translated genes (Table III). As mentioned above, class C is distinctive by the absence of a homologue of *orf11* and the presence of three unique genes (*orf14*, *orf15,* and *orf16*). When comparing the translated ORFs that are common to all classes, it is observed that the most conserved ones are at each end of the locus with ORFs 1, 2, and 13 sharing above 94% protein sequence identity between corresponding homologues. ORFs 3, 4, 8, 9, 10, and 12 share from 66 to 86% protein sequence identity, whereas the most divergent proteins are found in the middle of the locus with ORFs 5–7 sharing from 34 to 50% protein sequence identity.

*Gene Inactivation by the Deletion or Insertion of a Single Base (without Phase Variation)*—There are two glycosyltransferase genes that are found as inactive versions in some of the strains due to frameshift mutations. There is a missing A base at position 1,234 of *orf3* in four class A strains (ATCC 43432, ATCC 43446, OH4382, and OH4384). Based on BLAST searches, *orf3* was proposed to encode a 515-amino acid two-domain glycosyltransferase (The Sanger Center website address: www.sanger.ac.uk/Projects/C_jejuni/, predicted coding sequence Cj1135). The amino acid sequence at the N terminus (residues 1–250) is homologous to LgtF from *Haemophilus ducreyi* that encodes a β-1,4-glucosyltransferase that transfers glucose to heptose (28). The first domain of *orf3* is therefore the likely candidate for transferring the β-1,4-glucose to the inner heptose (Hep-I) in *C. jejuni*. The second domain (residues 250–515) of *orf3* is homologous to various glycosyltransferases, but it is not possible to deduce its specificity based on sequence homology alone. However, the frameshift mutation observed in four class A strains results in the expression of a 418-amino acid protein which means that the second domain is missing 98 residues. Because the four strains that have this frameshift mutation are also missing the β-1,2-glucose residue on the second heptose (Hep-II, see Fig. 1), we suggest that the second domain of *orf3* is a β-1,2-glucosyltransferase.

The second example of a glycosyltransferase gene that shows inactivation by frameshift mutation is *orf5* in strain OH4382 (missing A at base 71), *orf5-I* in ATCC 43449 (missing A at base 71), and *orf5-I* in ATCC 43456 (missing G at base 200). We reported previously that this gene encodes a β-1,4-*N*-acetylgalactosaminyltransferase and that its inactivation results in the expression of a truncated LOS in OH4382 (24). However, the inactivation of *orf5-I* in ATCC 43449 and ATCC 43456 does not result in LOS outer cores without GalNAc because these two strains have a second, functional, copy of this gene (*orf5-II*).

*Phase Variation Due to Homopolymeric G-tracts*—Four of the 11 *C. jejuni* strains lack G-tracts longer that 5 bases in their LOS biosynthesis locus (Fig. 2). Longer homopolymeric G-tracts are present in five LOS biosynthesis genes distributed among the seven other *C. jejuni* strains. Some of the G-tracts
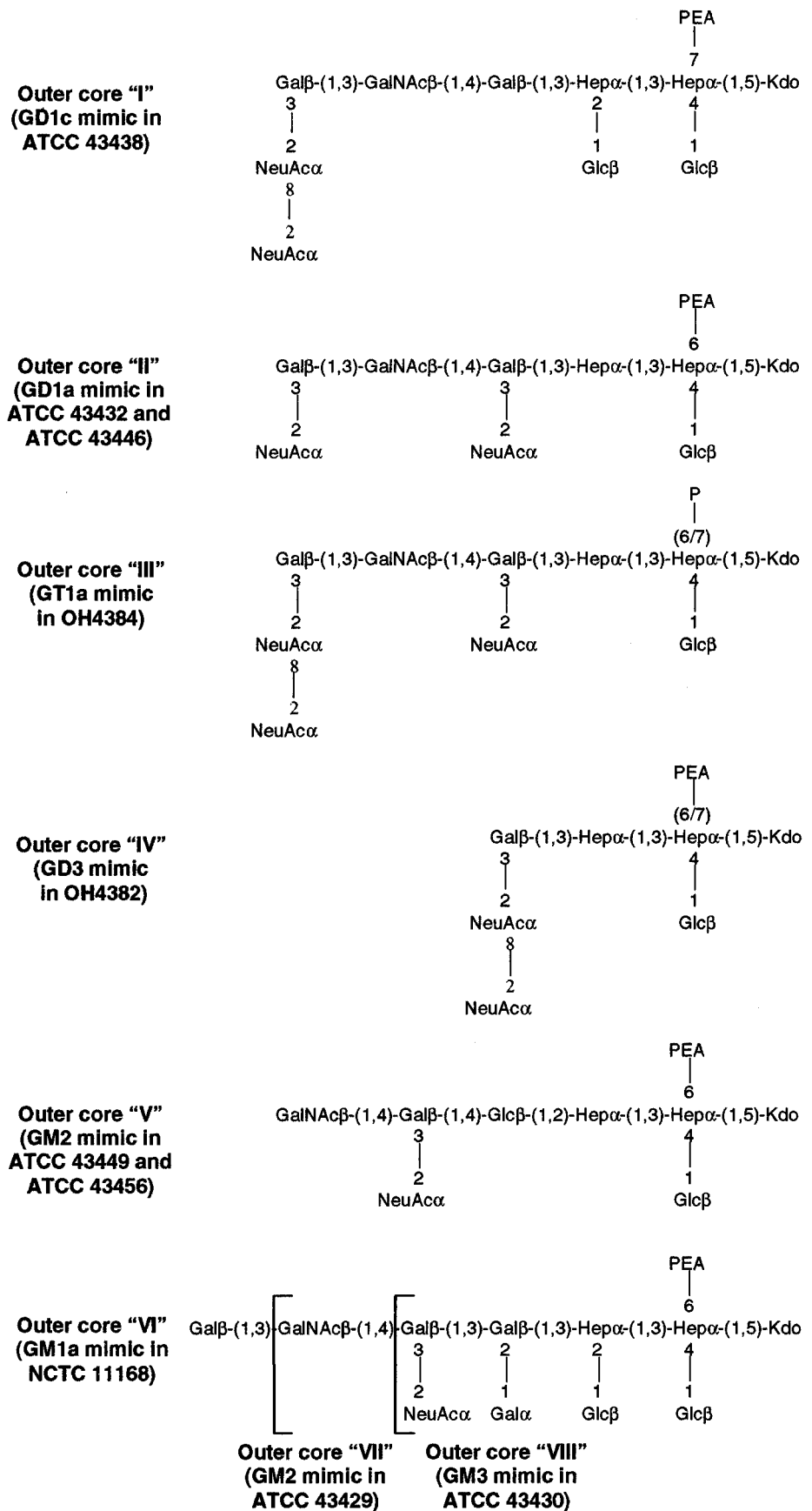
FIG. 1. **Outer regions of the LOS structures expressed by the *C. jejuni* strains studied in this work (see Table I for the references to the original publications of the structures).** The numbering system (I–VIII) is not based on any biological phenotype and serves only for referencing purposes. The LOS outer core structures display the oligosaccharide portions of the gangliosides indicated *below* the structure number.

**Outer core "I" (GD1c mimic in ATCC 43438)**

```
                                              PEA
                                               |
                                               7
Galβ-(1,3)-GalNAcβ-(1,4)-Galβ-(1,3)-Hepα-(1,3)-Hepα-(1,5)-Kdo
      3                              2         4
      |                              |         |
      2                              1         1
   NeuAcα                          Glcβ      Glcβ
      8
      |
      2
   NeuAcα
```

**Outer core "II" (GD1a mimic in ATCC 43432 and ATCC 43446)**

```
                                              PEA
                                               |
                                               6
Galβ-(1,3)-GalNAcβ-(1,4)-Galβ-(1,3)-Hepα-(1,3)-Hepα-(1,5)-Kdo
      3                       3               4
      |                       |               |
      2                       2               1
   NeuAcα                   NeuAcα          Glcβ
```

**Outer core "III" (GT1a mimic in OH4384)**

```
                                              P
                                               |
                                             (6/7)
Galβ-(1,3)-GalNAcβ-(1,4)-Galβ-(1,3)-Hepα-(1,3)-Hepα-(1,5)-Kdo
      3                       3               4
      |                       |               |
      2                       2               1
   NeuAcα                   NeuAcα          Glcβ
      8
      |
      2
   NeuAcα
```

**Outer core "IV" (GD3 mimic in OH4382)**

```
                                       PEA
                                        |
                                      (6/7)
         Galβ-(1,3)-Hepα-(1,3)-Hepα-(1,5)-Kdo
              3               4
              |               |
              2               1
           NeuAcα           Glcβ
              8
              |
              2
           NeuAcα
```

**Outer core "V" (GM2 mimic in ATCC 43449 and ATCC 43456)**

```
                                                    PEA
                                                     |
                                                     6
GalNAcβ-(1,4)-Galβ-(1,4)-Glcβ-(1,2)-Hepα-(1,3)-Hepα-(1,5)-Kdo
      3                                       4
      |                                       |
      2                                       1
   NeuAcα                                   Glcβ
```

**Outer core "VI" (GM1a mimic in NCTC 11168)**

```
                                                              PEA
                                                               |
                                                               6
Galβ-(1,3)-GalNAcβ-(1,4)-Galβ-(1,3)-Galβ-(1,3)-Hepα-(1,3)-Hepα-(1,5)-Kdo
                  3         2         2               4
                  |         |         |               |
                  2         1         1               1
               NeuAcα     Galα      Glcβ            Glcβ
```

**Outer core "VII" (GM2 mimic in ATCC 43429)**     **Outer core "VIII" (GM3 mimic in ATCC 43430)**

are unique to one gene and strain (such as in *orf7* of ATCC 43449), whereas some others are present in all representatives of a gene (such as in *orf16* in the three class C strains). Some of the G-tracts were found to be homogeneous, *i.e.* with a specific number of G bases that would result in the expression of a specific on or off phenotype. For instance, *orf6* (β-1,3-galacto-
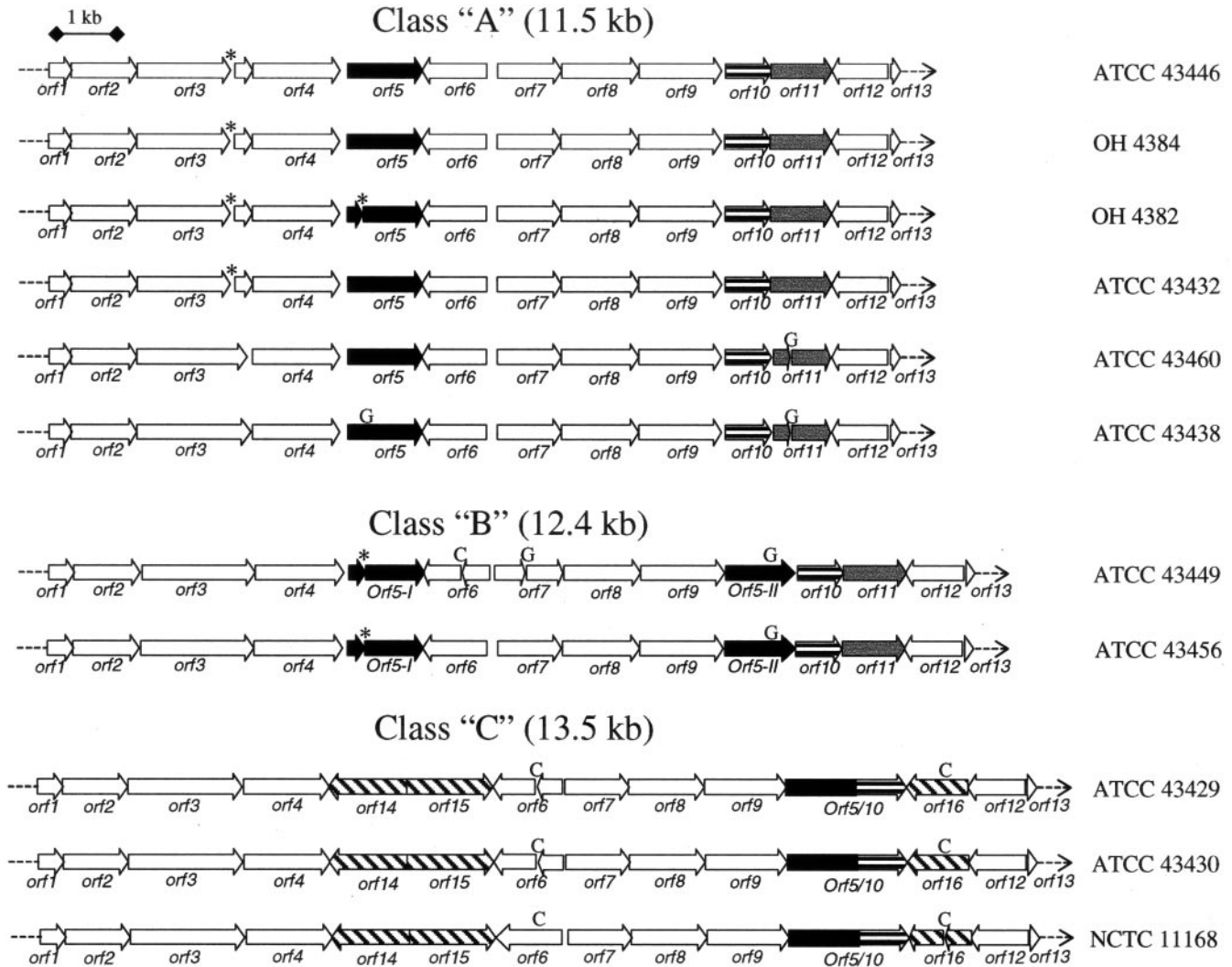
FIG. 2. **Genetic organization of the LOS biosynthesis loci of the different *C. jejuni* strains.** The proposed functions for the ORFs are described in Table II. The * indicates where a premature translation stop is observed for some of the strains. The *G* indicates where a poly(G) tract is observed (*C* shows where genes are translated on the complementary strands). Phase-variable ORFs that were observed to be mostly out of frame (see Table IV) are broken in *two arrows*. There is a gene (*orf11*, shown in *gray*), unique to classes A and B. Three genes (*orf14, orf15,* and *orf16*, shown with *downward stripes*) are unique to class C. The gene encoding the β-1,4-*N*-acetylgalactosaminyltransferase (*orf5*, shown in *black*) is found in one copy in class *A*, in two copies in class *B*, and as an in-frame fusion with the CMP-Neu5Ac synthetase (*orf10*, shown with *horizontal stripes*) in class C.

syltransferase) is inactive in ATCC 43429 and ATCC 43430 because it is found to have a homogeneous 9 G-tract that causes premature translation termination, consistent with the absence of a terminal β-1–3-Gal residue in these strains (Fig. 1, *structures VII* and *VIII*, and see Table I). Other G-tracts are heterogeneous with one of the variants being present more frequently. Determining the proportions of each variant was found to be difficult because heterogeneity was sometimes observed even when chromosomal DNA was isolated from single colonies. We defined the "most frequent variant" as the one corresponding to the strongest signal on a DNA sequencing electropherogram when we sequenced a PCR product obtained using as template chromosomal DNA isolated from a confluent plate. Because NCTC 11168 was sequenced from a plasmid library, specific numbers were reported for each variant of *orf6* and *orf16* for this strain (see Table IV for references). In the case of *orf6* from NCTC 11168, the most frequent variant has 8 G (in-frame) which is consistent with the LOS outer core (structure VI) having a terminal β-1,3-Gal residue (Ref. 22 and see Fig. 1). However, it is not always possible to correlate the most frequent variants with the published structures. For instance,

the LOS outer core structure of ATCC 43449 was reported to be sialylated (Fig. 1, *structure V*, and see Table I), but it contains an α-2,3/2,8-sialyltransferase gene (*orf7*) mostly as an out-of-frame variant (Table IV). It is possible that the level of active *orf7* in ATCC 43449 is sufficient to produce LOS with enough of the sialic acid residue for it to be detected by chemical analysis. Because the phase-variable genes are heterogeneous, it is also probable that the proportion of active/inactive variants will vary between laboratories depending on the number of passages of the strain and whether practices such as sub-culturing from isolated colonies are used or not. We avoided sub-culturing from single colonies because our original stocks were not single colonies and to avoid enriching specific variants.

*Single Mutations Leading to the Inactivation of a Glycosyltransferase*—There are only eight base differences between the LOS biosynthesis loci of *C. jejuni* NCTC11168 and ATCC 43430 although they express different LOS outer cores (Fig. 1, *structures VI* and *VIII*, respectively). Six of these base differences cause frameshift changes in *orf6* (β-1,3-galactosyltransferase) and in *orf16* (unknown function, 5 bases are missing in NCTC 11168). One of the base differences causes a silent mutation in

TABLE II

*Proposed functions for the LOS biosynthesis locus ORFs (see Fig. 2 for position) and summary of the sequence comparison and experimental evidence*

| ORF | Gene | Proposed function | Evidence by sequence homology and/or experimental evidence |
|---|---|---|---|
| 1 | Cj1133[a] waaC | Heptosyltransferase I | Homology with RfaC and complementation (36) |
| 2 | Cj1134 htrB | Lipid A biosynthesis acyltransferase | Homology with WaaM |
| 3 | Cj1135 | Two-domain glucosyltransferase: β-1,4-glucosyltransferase (N-terminal domain) and β-1,2-glucosyltransferase (C-terminal domain) | The N-terminal domain is homologous with LgtF and the C-terminal domain with various glycosyltransferases. Premature translation stop in the C terminus of all the strains that don't have a β-1,2-glucose residue[b] |
| 4 | Cj1136 | β-1,3-Galactosyltransferase | Homology with various galactosyltransferases |
| 5 | Cj1143 cgtA | β-1,4-N-Acetylgalactosaminyltransferase (to Gal) | *In vitro* activity (24) and knock-out mutant (25) |
| 6 | Cj1139c cgtB | β-1,3-Galactosyltransferase (to GalNAc) | Homology with various galactosyltransferases *In vitro* activity (24) and knock-out mutant (22) |
| 7 | Cj1140 cst-II cst-III | α-2,3 or α-2,3/α-2,8-sialyltransferase (to Gal/Neu5Ac) | *In vitro* activity (24) and knock-out mutant (25) |
| 8 | Cj1141 neuB1 | Sialic acid synthase | Homology with various sialic acid synthases and complementation (23) |
| 9 | Cj1142 neuC1 | N-Acetylglucosamine-6-phosphate 2-epimerase | Homology with various N-acetylglucosamine-6-phosphate 2-epimerase and knock-out mutant (25) |
| 10 | Cj1143 | CMP-Neu5Ac synthetase | Homology with various CMP-Neu5Ac synthetase and *in vitro* activity[b] |
| 11 | | Putative acetyltransferase | Homology with various acetyltransferases |
| 12 | Cj1146c | Putative glycosyltransferase | Homology with various glycosyltransferases |
| 13 | Cj1148 waaF | Heptosyltransferase II | Homology with RfaF |
| 14 | Cj1137c | Putative glycosyltransferase | Homology with various glycosyltransferases |
| 15 | Cj1138 | Putative glycosyltransferase | Homology with various glycosyltransferases |
| 16 | Cj1145c | Putative ORF | |

[a] The Cj*XXXX* numbers correspond to the gene numbering of the "genome" strain (NCTC 11168, www.sanger.ac.uk/Projects/C_jejuni/). Other gene nomenclature found in the literature is also indicated.

[b] This work.

TABLE III

*Comparison of the deduced amino acid sequences of the ORFs (see Fig. 2 for position) of the three classes of LOS biosynthesis locus*

| ORF | Proposed function | Identity of the deduced polypeptides | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Class A | Class B | Class C | Classes A and B | Classes A and C | Classes B and C | Classes A, B, and C |
| | | % | | | | | | |
| 1 | Heptosyltransferase I[a] | 97 | 97 | 98 | 96 | 97 | 96 | 95 |
| 2 | Lipid A biosynthesis acyltransferase | 95 | 98 | 99 | 95 | 95 | 96 | 94 |
| 3 | Two-domain glucosyltransferase | 88[b] | 99 | 100 | 84[b] | 84[b] | 88 | 84[b] |
| 4 | Putative galactosyltransferase | 94 | 100[b] | 100 | 94[b] | 74 | 75[b] | 72[b] |
| 5 | β-1,4-N-Acetylgalactosaminyltransferase | 73 | 100 | 99[c] | 38[d] | 42[c] | 64[c] | 34[c] |
| 6 | β-1,3-Galactosyltransferase[d] | 59 | —[e] | 100[b] | 59 | 47[b] | 58[b] | 47[b] |
| 7 | Sialyltransferase | 95 | 94 | 100 | 92 | 51 | 50 | 50 |
| 8 | Sialic acid synthase | 97 | 99 | 100 | 97 | 76 | 76 | 76 |
| 9 | N-Acetylglucosamine-6-phosphate 2-epimerase | 97 | 99 | 100 | 95 | 68 | 69 | 67 |
| 10 | CMP-Neu5Ac synthetase | 96 | 100 | 100[f] | 95 | 67[f] | 69 | 66[f] |
| 11 | Acetyltransferase | 95 | 99 | NA[g] | 94 | NA | NA | NA |
| 12 | Putative glycosyltransferase | 95 | 99[b] | 100 | 91[b] | 86 | 89[b] | 86[b] |
| 13 | Heptosyltransferase II[h] | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 14 | Putative glycosyltransferase | NA | NA | 100 | NA | NA | NA | NA |
| 15 | Putative glycosyltransferase | NA | NA | 100 | NA | NA | NA | NA |
| 16 | Hypothetical protein | NA | NA | 100[b] | NA | NA | NA | NA |

[a] The heptosyltransferase I alignments were done with the C-terminal sequences only (115 residues).

[b] These alignments contain one or more sequences that have premature translation stops. No "penalty" was calculated for the truncations, *i.e.* the % identity is calculated using only the truncated coding regions.

[c] For this alignment we used only the N-terminal domain (β-1,4-N-acetylgalactosaminyltransferase) of the fusion ORF 5/10 from the Class C.

[d] The alignments that showed less than 70% conserved residues are underlined to highlight the most divergent genes.

[e] The *cgtB* gene from *C. jejuni* O:23 contains multiple frameshift mutations and was not included in any of the alignments.

[f] For this alignment we used only the C-terminal domain (CMP-sialic acid synthetase) of the fusion ORF 5/10 from the Class C.

[g] NA, not applicable (no homologue in one of the classes).

[h] The heptosyltransferase II alignments were done with the N-terminal sequences only (40 residues).

*orf16*, whereas the last base difference causes an amino acid change (Cys-92 → Tyr, NCTC 11168 → ATCC 43430) in *orf5/10* (β-1,4-N-acetylgalactosaminyltransferase/CMP-NeuAc synthetase natural fusion). Because the LOS outer core of ATCC 43430 is truncated at the second inner Gal residue (Fig. 1, *structure VIII*), we suspected that this mutation was respon-

sible for the inactivation of the β-1,4-N-acetyl galactosaminyltransferase in ATCC 43430. We cloned *orf5/10* from both NCTC 11168 and ATCC 43430 and expressed them in *E. coli*. We found that both versions have similar CMP-NeuAc synthetase activity, whereas only the NCTC11168 version has β-1,4-N-acetylgalactosaminyltransferase activity (Table V).

TABLE IV

*Location of homopolymeric G-tracts in the LOS biosynthesis locus of the various C. jejuni strains*

| Strain (class) | Gene | Proposed function | G-tract length | Translation of the most frequent variant |
|---|---|---|---|---|
| ATCC 43438 (A) | *orf5* | β-1,4-GalNAc transferase | 9 | In-frame |
| ATCC 43456 (B) | *orf5-II* | β-1,4-GalNAc transferase | 9/**10**[a] | In-frame |
| ATCC 43449 (B) | *orf5-II* | β-1,4-GalNAc transferase | 9/**10** | In-frame |
| ATCC 43449 (B) | *orf6*[b] | β-1,3-Gal transferase | 8/**9** | Out of frame |
| ATCC 43429 (C) | *orf6*[b] | β-1,3-Gal transferase | 9 | Out of frame |
| ATCC 43430 (C) | *orf6*[b] | β-1,3-Gal transferase | 9 | Out of frame |
| NCTC11168 (C) | *orf6*[b] | β-1,3-Gal transferase | **8**/9[c] | In-frame |
| ATCC 43449 (B) | *orf7* | α-2,3-/α-2,8-sialyltransferase | 9/**10** | Out of frame |
| ATCC 43460 (A) | *orf11* | Putative acetyltransferase | 8 | Out of frame |
| ATCC 43438 (A) | *orf11* | Putative acetyltransferase | 10/**11** | Out of frame |
| ATCC 43429 (C) | *orf16*[b] | unknown | **8**/9[d] | In-frame |
| ATCC 43430 (C) | *orf16*[b] | unknown | **8**/9[d] | In-frame |
| NCTC11168 (C) | *orf16*[b] | unknown | **8**/9[e] | Out of frame |

[a] The number in bold indicates the most frequent variant when heterogeneity was observed. We defined the "most frequent variant" as the one corresponding to strongest signal on a DNA sequencing electropherogram when we sequenced a PCR product obtained using as template chromosomal DNA isolated from a confluent plate.

[b] *orf6* and *orf16* are translated in the opposite orientation of the other ORFs with G-tracts.

[c] Linton *et al.* 2000 (22).

[d] There are four additional As upstream of the G-tract in *orf16* of ATCC 43429 and 43430.

[e] The Sanger Centre website is www.sanger.ac.uk/Projects/C_jejuni/.

TABLE V

*Enzymatic assays of various alleles of the β-1,4-N-acetylgalactosaminyltransferase (orf5, orf5-II, and orf5/10) expressed as fusions with MalE in E. coli*

| Gene | *C. jejuni* strain from which *cgtA* was cloned | CMP-Neu5Ac synthetase[a] | β-1,4-*N*-Acetylgalactosaminyltransferase[b] | | |
|---|---|---|---|---|---|
| | | | **Gal**-β1,4-Glc- | Neu5Acα2–3-**Gal**-β1,4-Glc- | Neu5Acα2–3-**Gal**-β1,4-Glc-<br>\|<br>Neu5Acα2-8- |
| | | | *specific activity*[c] | | |
| *orf5* | ATCC 43438 (A) | 0 | 757 | 0 | 0 |
| *orf5* | OH4384 (A) | 0 | 20 | 2,797 | 0 |
| *orf5-II* | ATCC 43456 (B) | 0 | 0 | 137,790 | 134,810 |
| *orf5/10* | ATCC 43430 (C) | 317,230 | 0 | 0 | 0 |
| *orf5/10* | NCTC 11168 (C) | 517,260 | 0 | 8,123 | 7,357 |

[a] The CMP-Neu5Ac synthetase was assayed using CTP, Neu5Ac, Gal-β1,4-GlcNAc-FCHASE, and a purified α-2,3-sialyltransferase (NST-27) in a coupled assay that measured the production of Neu5Acα-2,3-Gal-β1,4-GlcNAc-FCHASE.

[b] The β-1,4-*N*-acetylgalactosaminyltransferase activity was assayed using UDP-GalNAc as donor and fluorescein-labeled (-FCHASE) oligosaccharides. The GalNAc residue is transferred to the Gal residue (in bold).

[c] The specific activity is expressed in microunits (picomoles of product/min)/mg of total protein in the extract. We report the means of triplicate experiments.

*Mutations Leading to Glycosyltransferases with Different Glycan Acceptor Specificities*—Although the β-1,4-*N*-acetylgalactosaminyltransferase alleles from the three classes are clearly homologous, the level of conservation among them is only 34% (Table III). We expressed representatives from each class as C-terminal fusions with the maltose-binding protein in *E. coli*. The acceptor preference was found to vary significantly (Table V) with the ATCC 43438 version using only a nonsialylated acceptor, the version from OH4384 using only a mono-sialylated acceptor, and the versions from ATCC 43456 and NCTC 11168 are able to use both a mono-sialylated and a di-sialylated acceptors. In most cases the acceptor specificity correlates with the natural acceptor because only ATCC 43438 has no sialic acid on the inner Gal residue of the LOS outer core, and the three other strains (OH4384, ATCC 43456, and NCTC 11168) have a single sialic acid on the inner Gal residue (Fig. 1). The ability of the β-1,4-*N*-acetylgalactosaminyltransferase from ATCC 43456 and NCTC 11168 to use a di-sialylated acceptor could seem superfluous because these strains express LOS outer cores with a single sialic acid on the inner Gal residue. However, Prendergast and Moran (29) recently reported a *C. jejuni* outer core mimicking GD2 (*i.e.* GalNAc-β-1,4-[Neu5Ac-α-2,8-Neu5Ac-α2,3-]-Gal-inner core), which suggests that this strain could contain a CgtA version that is related to the one found in either ATCC 43456 or NCTC 11168. In addition some *C. jejuni* strains were reported to express a GQ1b epitope in their LOS outer core based on antibody prob-

ing (30). Although the expression of an "authentic" GQ1b mimic requires confirmation by structural analysis, the ability of the β-1,4-*N*-acetylgalactosaminyltransferase from ATCC 43456 and NCTC 11168 to use a di-sialylated acceptor suggests that this structure could exist in the outer core of some *C. jejuni* strains.

Another example of mutations leading to different acceptor specificities is provided by *orf7*, which was named *cst-II* when we cloned it from *C. jejuni* OH4384 (24). We will use this designation for all of the versions from classes A and B. Gerry *et al.* (25) showed that *orf7* from ATCC 43429 is responsible for transferring the α-2,3-sialic acid and named this gene *cst-III*, a designation that we will use for class C *orf7*. An alignment of the deduced protein sequences of the *orf7* (sialyltransferase) versions from all the classes gave 50% identity. However, when the classes A and B versions are aligned together, the level of protein sequence identity rises to 92% (Fig. 3), whereas the three class C versions are 100% identical between themselves. Pairwise alignments between Cst-III and each variant of Cst-II gave 52% protein sequence identity on average.

Because Cst-II from OH4382 and OH4384 are identical, there are seven distinct Cst-II amino acid sequences (Fig. 3). We cloned and expressed six of them in *E. coli* and assayed the recombinant Cst-IIs for α-2,3-sialyltransferase and α-2,8-sialyltransferase activities (Table VI) using Gal-β-1,4-Glc-FCHASE and Neu5Ac-α-2,3-Gal-β-1,4-Glc-FCHASE, respectively, as acceptors. We found four versions (OH4382/84, ATCC

```
                   10        20        30        40        50        60        70        80        90
                    |         |         |         |         |         |         |         |         |
43446      ......A.........................................C.A...T.NF.................T.......Y.......E.
43456      ......A.........................................C.T...T.NF.................T.......Y.......E.
43432      ......A.........................................C.A...T.GF.................T.......Y.......E.
43438      ......S.........................................F.A...N.GL.................T.......Y.......E.
OH4382/84  ......A.........................................C.A...N.IL.................T.......Y.......E.
43449      ......S.........................................C.A...N.SL.................I.......Y.......E.
43460      ......S.........................................C.A...N.SL.................T.......F.......Q.
           ****** ****************************************** * *** *  ***************** ******* ****** *
Consensus  MKKVIIAGNGPSLKEIDYSRLPNDFDVFRCNQFYFEDKYYLGKKCKAVFYNPXLFFEQYYTLKHLIQNQEYETELIMCSNYNQAHLENEN


                   100       110       120       130       140       150       160       170       180
                    |         |         |         |         |         |         |         |         |
43446      ........................E.........F..........A.....................Q..SS.......E.......D...
43456      ........................E.........F..........A.....................Q..SS.......E.......D...
43432      ........................E.........F..........A.....................Q..SS.......E.......D...
43438      ........................E.........L..........A.....................Q..SS.......E.......D...
OH4382/84  ........................D.........F..........A.....................Q..SS.......K.......N...
43449      ........................E.........F..........T.....................N..GG.......K.......N...
43460      ........................E.........F..........T.....................Q..SS.......K.......N...
           ************************ ********* ********** ***************** ** ******* ****** ******* ***
Consensus  FVKTFYDYFPDAHLGYDFFKQLKEFNAYFKFHEIYFNQRITSGVYMCAVAIALGYKEIYLSGIDFYQNGSSYAFDTKQENLLKLAPDFKN


                   190       200       210       220       230       240       250       260       270
                    |         |         |         |         |         |         |         |         |
43446      .R......................K...............................................S..............V.
43456      .R......................K...............................................S..............V.
43432      .R......................K...............................................S..............V.
43438      .R......................K...............................................S..............I.
OH4382/84  .N......................K...............................................S..............I.
43449      .N......................E...............................................T..............I.
43460      .N......................E...............................................T..............I.
           * ********************** ******************************************* ************* *
Consensus  DRSHYIGHSKNTDIKALEFLEKTYKIKLYCLCPNSLLANFIELAPNLNSNFIIQEKNNYTKDILIPSSEAYGKFSKNINFKKIKIKENIY


                   280       290
                    |         |
43446      ........R............
43456      ........R............
43432      ........R............
43438      ........R............
OH4382/84  ........R............
43449      ........K............
43460      ........R............
           ******* ************
Consensus  YKLIKDLLRLPSDIKHYFKGK
```
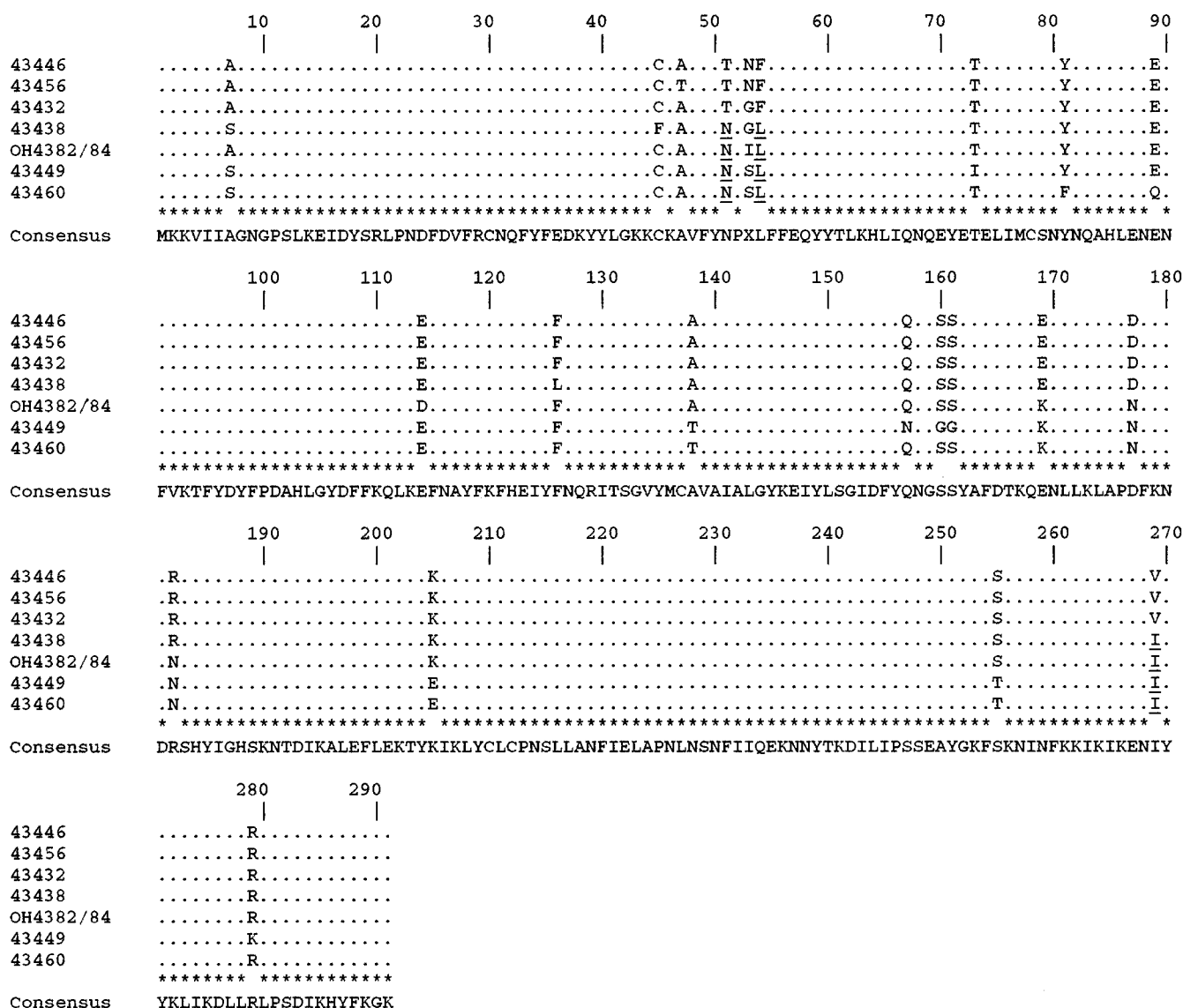
FIG. 3. **Alignment of the *cst-II* sialyltransferase (*orf7*) deduced amino acid sequences.** The *C. jejuni* strain numbers are indicated to the *left* of the sequences, and the GenBank™ accession numbers are indicated in Table I. The three residues (Asn-51, Leu-54, and Ile-269) that are specific for the bi-functional Cst-II variants are *underlined*.

43438, ATCC 43449, and ATCC 43460) that are bi-functional (both α-2,3- and α-2,8-sialyltransferase activities), and two versions (ATCC 43432 and ATCC 43446) that have only the α-2,3-sialyltransferase activity (Table VI). An alignment of the amino acid sequences of the various Cst-II versions (Fig. 3) indicated that only three residues (Asn-51, Leu-54, and Ile-269) were specific for the bi-functional Cst-II versions. We used site-directed mutagenesis to determine which of these residues are essential for bi-functional sialyltransferase activity. An Asn-51 → Thr substitution in Cst-II from OH4384 completely abolished the α-2,8-sialyltransferase activity (Table VI). The opposite substitution (Thr-51 → Asn) in the mono-functional Cst-II from ATCC 43446 conferred it the ability to perform both activities (α-2,3- and α-2,8-sialyltransferase). The other two residues (Leu-54 and Ile-269) unique to bi-functional Cst-II variants as well as the very variable residue 53 were found to affect the relative ratios of α-2,3- and α-2,8-sialyltransferase activities (Table VI and data not shown), but only Asn-51 was found to be absolutely essential for α-2,8-sialyltransferase activity. Although the *in vitro* assays with the various recombinant Cst-IIs allowed us to determine which versions are mono- or bi-functional, the levels of activities vary considerably be-

tween the various versions (Table VI). SDS-PAGE analyses indicated that all the versions were expressed at similar levels (data not shown). Two Cst-II versions (ATCC 43449 and ATCC 43460) have low α-2,3-sialyltransferase activity, whereas the Cst-II from OH4384 has both low α-2,3- and low α-2,8-sialyltransferase activities (Table VI). The amino acid substitution Ile-53 → Gly increased both activities of the OH4384 version (Table VI) which suggests that this residue has an important impact on the level of *in vitro* activity. It is also noticeable that the two versions (ATCC 43449 and ATCC 43460) that have much lower α-2,3- than α-2,8-sialyltransferase activity both have the same residue (a serine) at position 53 (Fig. 3).

DISCUSSION

Genetic loci involved in the biosynthesis of cell-surface carbohydrates have been identified in many bacteria as a result of the sequencing of entire genomes. However, only a few studies have looked at the corresponding loci of strains expressing distinct carbohydrate structures. Different gene complements and phase variation due to homopolymeric G-tracts were shown to be involved in the variability of LOS outer core structures in *N. meningitidis* (31). Different gene complements

TABLE VI

*Spectrum of α-2,3- and α-2,8 sialyltransferase activities of the recombinant Cst-II variants*

All the constructs were expressed as MalE/Cst-II fusion in *E. coli*.

| *C. jejuni* strain from which *cst-II* was cloned | Sialyltransferase activity | |
|---|---|---|
| | α-2,3-Sialyltransferase[a] | α-2,8-Sialyltransferase[b] |
| | *specific activity[c]* | |
| ATCC 43432 | 8,757 | 0 |
| ATCC 43446 | 8,197 | 0 |
| ATCC 43446 (Thr 51 → Asn)[d] | 9,600 | 57 |
| OH4382/84[e] | 133 | 177 |
| OH4382/84 (Asn 51 → Thr) | 87 | 0 |
| OH4382/84 (Ile 53 → Gly) | 6,250 | 12,407 |
| ATCC 43438 | 10,227 | 7,290 |
| ATCC 43449 | 50 | 1,943 |
| ATCC 43460 | 100 | 1,437 |

[a] The α-2,3-sialyltransferase activity was assayed using CMP-Neu5Ac as donor and **Gal**β-**1**,4-Glc-FCHASE as acceptor. The sialic acid (Neu5Ac) residue is transferred to the Gal residue (in bold).

[b] The α-2,8-sialyltransferase activity was assayed using CMP-Neu5Ac as donor and **Neu5Ac**α-2–3-Galβ-1,4-Glc-FCHASE as acceptor. The sialic acid (Neu5Ac) residue is transferred to the Neu5Ac residue (in bold).

[c] The activity is expressed in microunits (picomoles of product/min)/mg of total protein in the extract. We report the means of triplicate experiments.

[d] Mutations are indicated in parentheses.

[e] Cst-II from *C. jejuni* OH4382 and OH4384 are identical.

were also observed in the corresponding loci responsible for various inner core structures in *E. coli* (32). Comparative genomics studies of the capsular polysaccharide biosynthesis loci from *Streptococcus pneumoniae* strains of different serotypes have shown evidence of recombination events resulting in different gene complements (33–34). Corresponding glycosyltransferase genes that had diverged were also proposed to contribute to the capsular variability by transferring the same sugar unit to create different linkages, although no biochemical data were reported to support the proposed functions (33).

The presence of a large number of ganglioside mimics in various *C. jejuni* strains prompted us to investigate the genetic basis for this variation. The general organization of the various LOS biosynthesis loci allowed them to be grouped in three classes and demonstrated that not all of the differences in LOS outer cores are due to the different gene complements. Previous work had also shown that phase variation using homopolymeric G tracts (22) and gene inactivation by the deletion or insertion of a single base (without phase variation) were also responsible for some of the variations in LOS outer core structures (24). By combining comparative genomics of LOS biosynthesis loci from strains expressing different LOS outer cores with functional assays, it was possible to determine that *C. jejuni* also uses glycosyltransferase alleles to produce enzymes that are inactive or that show different acceptor specificities. We propose that each of the differences in the structure of LOS outer cores displayed by the 11 different *C. jejuni* strains can be explained by either one or more of the five genetic mechanisms we have described. Transcriptional regulation was not examined in this study, and it is possible that the expression of some glycosyltransferases (or of other carbohydrate biosynthesis enzymes) would be induced or repressed under varying growth conditions or during infection. However this would not change the potential of a strain with specific glycosyltransferase alleles to make the LOS outer core structure(s). Because some of the variations are due to amino acid substitutions or frameshift mutations, the regulation of the DNA repair system is also likely to have an impact on the possibility of strain to vary its LOS outer core.

Class B is clearly an evolutionary intermediate between classes A and C as it seems to have evolved from a class A locus by duplication of *orf5* (into *orf5-I* and *orf-II*, see Fig. 2). At least two more recombination events would have been necessary to generate a class C locus by the insertion of *orf14*, *orf15,* and *orf16* and the deletion of *orf5-I* and *orf11*. The three class C loci

studied here also have *orf5-II* and *orf10* as an in-frame fusion. It is certainly possible that other evolutionary intermediates exist with different combinations of inserted/deleted ORFs and with *orf5-II* and *orf10* either as separate ORFs or as an in-frame fusion. Although class C loci have three unique genes, there seems to be a need for only two additional glycosyltransferases when the outer core structures are compared. Class C outer cores have two linkages (a Gal-β1,3-Gal and a Gal-α1,2-Gal, see Fig. 1) that are not present in classes A and B outer cores. The *orf14* and *orf15* both show homology with various glycosyltransferases (data not shown) and would be good candidates to make these two linkages although current experimental evidence is not sufficient to confirm these assignments. The *orf16* is a hypothetical ORF with no homologue in GenBank™, and it is not possible to determine its role, if any, in LOS biosynthesis. In families A and B, *orf11* has no clear function in LOS biosynthesis. It shows homology with various acetyltransferases (data not shown), but acetylation of the *C. jejuni* LOS structures was not reported, although it could have been overlooked.

The divergence observed between ATCC 43438 and the other class A loci in the region from nt 4500 to 5700 is interesting from both the functional and evolutionary aspects. In this region, ATCC 43438 is much more similar to ATCC 43449 from class B than to the other class A loci. The other class A loci are themselves closer to ATCC 43456 (the other class B locus) in the corresponding region. Consequently, ATCC 43438 could either be an evolutionary intermediate between classes A and B, and ATCC 43456 would have acquired the class A corresponding region by lateral gene transfer. The opposite sequence of events is as likely, *i.e.* ATCC 43456 would be an intermediate between classes A and B loci and ATCC 43438 would have acquired the region from nt 4500 to 5700 by lateral gene transfer.

It is also noteworthy that the region from nt 4500 to 5700 spans both the *orf5* and *orf6* that encode a β-1,4-*N*-acetylgalactosaminyltransferase and a β-1,3-galactosyltransferase, respectively. Because *orf5* and *orf6* are translated in opposite orientations, the divergence of the region from nt 4500 to 5700 results in the large number of amino acid substitutions observed in the C terminus of both the β-1,4-*N*-acetylgalactosaminyltransferase and the β-1,3-galactosyltransferase of ATCC 43438 when they are compared with the corresponding glycosyltransferases from the other loci (data not shown). These two genes seem to have evolved to accommodate the presence of a

nonsialylated acceptor in the inner core of *C. jejuni* ATCC 43438. A functional assay of the recombinant β-1,4-*N*-acetyl-galactosaminyltransferase from ATCC 43438 confirmed that it is specific for a nonsialylated acceptor (Table V).

The absence or presence of activity of a specific glycosyl-transferase can also have an impact on the activity of other glycosyltransferases. Based on the examination of the LOS outer core structures, Nam Shin *et al.* (15) have suggested that the presence of a β-1,2-glucosyl residue on Hep-II would prevent sialylation of the inner β-1,3-galactosyl residue, possibly because of steric hindrance. Consequently the inactivation of the second domain of *orf3* by a frameshift mutation results in both the absence of a β-1,2-glucosyl residue on Hep-II and makes possible the sialylation of the inner β-1,3-Gal residue by Cst-II (*orf7*). We suggested previously (24) that the inner sialic acid was added by the product of *cst-I*, a gene that was cloned from *C. jejuni* OH4384 by activity screening and found downstream of the *prfB* gene (Cj1455), *i.e.* outside of the LOS biosynthesis locus. However, *cst-I* was shown to be absent from some strains that have a sialic acid on the inner β-1,3-Gal residue of their LOS outer core (data not shown), and consequently it is unlikely to be responsible for LOS sialylation.

We reported previously (24) that *orf7* from *C. jejuni* ATCC 43446 (the O:19 type strain) encoded an α-2,3-sialyltransferase, whereas the version from *C. jejuni* OH4382/84 had both α-2,3- and α-2,8-sialyltransferase activities. We named these two versions mono-functional Cst-II and bi-functional Cst-II, respectively. Guerry *et al.* (25) showed that the corresponding gene in ATCC 43429 is responsible for sialylation of the LOS outer core and named it Cst-III because it only showed 53% protein sequence identity with Cst-II from *C. jejuni* OH4384. When extending the comparison of *orf7* to the other strains, it appears that classes A and B all have slightly divergent Cst-II versions, whereas class C strains all have an identical version of Cst-III. We showed that one of the variable residues among the Cst-II versions results in either a mono-functional Cst-II (Thr-51) or a bi-functional Cst-II (Asn-51). Although Cst-III also has Asn-51, it seems to have only α-2,3-sialyltransferase activity (mono-functional) as observed in the LOS outer core structures (Fig. 1) and from *in vitro* assays (data not shown). Because Cst-II and Cst-III have diverged significantly, it is not too surprising that the presence of Asn-51 in Cst-III is not sufficient to confer it α-2,8-sialyltransferase activity. The low level of protein sequence conservation between Cst-II and Cst-III might be a consequence of adaptation to significantly different acceptor environments: in classes A and B the acceptor β-1,3-Gal residue is next to an un-substituted sugar residue (either GalNAc, Hep, or Glc), whereas in class C the acceptor β-1,3-Gal is attached to a Gal residue that is substituted with an α-1,2-Gal residue (Fig. 1).

The *in vitro* assays allowed us to determine which Cst-IIs are mono- or bi-functional, although the levels of activities vary considerably between the Cst-II versions. Comparison of the amino acid sequences and site-directed mutagenesis suggested that some residues (such as a glycine at position 53) have a large impact on the level of *in vitro* activities. It is unclear if these residues affect the stability of the recombinant enzyme or the efficiency of catalysis. Because the wild type Cst-II versions are known to be functional in their respective strains, it is probable that the less active versions are still active enough *in vivo* to carry efficient LOS sialylation.

The availability of variable cell-surface structures should be advantageous to a pathogen in order to evade the immune system. The five different modulation mechanisms described in this work can be effective over various time scales. The different gene complements are probably a result of evolution as well

as of lateral gene transfers. There are at least two other distinct classes of LOS biosynthesis loci based on the sequences reported for *C. jejuni* LIO87 (GenBank™ accession number AF400669), *C. jejuni* ATCC 43431 (GenBank™ accession number AF411225), and *C. jejuni* 81116 (GenBank™ accession numbers AF343914 and AJ131360). These LOS loci were not included in this study because the corresponding LOS outer core structures have not been determined for *C. jejuni* LIO87 and 81116, whereas the LOS outer core of *C. jejuni* ATCC 43431 does not mimic ganglioside structures (14). Nevertheless these types of LOS loci certainly expand the pool of genes that could be recombined in the LOS biosynthesis loci of *C. jejuni*.

Although no study has shown directly that phase variation provides an advantage during the course of *C. jejuni* infection, the high level of heterogeneity of some of the homopolymeric G-tracts suggests that a mixture of LOS outer core structures is likely to be expressed in many cases.[3] Although some of the frameshift mutations were found to be more stable than the homopolymeric G-tracts, the "one-base difference" between *C. jejuni* OH4382 and OH4384 has occurred during the course of an infection because these two strains were isolated from siblings (35). Because a one amino acid substitution can change Cst-II from a mono- to bi-functional sialyltransferase (and vice versa), it is also possible that such mutations could occur during the course of an infection or an outbreak.

In this work we have shown that *C. jejuni* can use up to five mechanisms to vary its LOS outer cores. These mechanisms can involve as little as a 1-base or a one amino acid change or be more substantive, as in the acquisition of new genes. It will be interesting to determine whether the expression of other *C. jejuni* cell-surface carbohydrates involves as many different regulatory and modulating mechanisms or if other pathogens have the same repertoire of modulating mechanisms.

REFERENCES

1. Roberts, I. S. (1996) *Annu. Rev. Microbiol.* **50,** 285–315
2. Preston, A., Mandrell, R. E., Gibson, B. W., and Apicella, M. A. (1996) *Crit. Rev. Microbiol.* **22,** 139–180
3. Castric, P., Cassels, F. J., and Carlson, R. W. (2001) *J. Biol. Chem.* **276,** 26479–26485
4. Moxon, E. R., and Kroll, J. S. (1990) *Curr. Top. Microbiol. Immunol.* **150,** 65–85
5. Allos, B. M. (2001) *Clin. Infect. Dis.* **32,** 1201–1206
6. Bacon, D. J., Szymanski, C. M., Burr, D. H., Silver, R. P., Alm, R. A., and Guerry, P. (2001) *Mol. Microbiol.* **40,** 769–777
7. Guerry, P., Szymanski, C. M., Prendergast, M. M., Hickey, T. E., Ewing, C. P., Pattarini, D. L., and Moran A. P. (2002) *Infect. Immun.,* in press
8. Jacobs, B. C., Rothbarth, P. H., van der Meché, F. G. A., Herbrink, P., Schmitz, P. I. M., de Klerk, M. A., and van Doorn, P. A. (1998) *Neurology* **51,** 1110–1115
9. Moran, A. P., Prendergast, M. M., and Appelmelk, B. J. (1996) *FEMS Immunol. Med. Microbiol.* **16,** 105–115
10. Yuki, N., Handa, S., Taki, T., Kasama, T., Takahashi, M., Saito, K., and Miyatake, T. (1992) *Biomed. Res.* **13,** 451–453
11. Aspinall, G. O., McDonald, A. G., Raju, T. S., Pang, H., Moran, A. P., and Penner, J. L. (1993) *Eur. J. Biochem.* **213,** 1017–1027
12. Aspinall, G. O., McDonald, A. G., Raju, T. S., Pang, H., Kurjanczyk, L. A., Penner, J. L., and Moran, A. P. (1993) *Eur. J. Biochem.* **213,** 1029–1037
13. Aspinall, G. O., McDonald, A. G., Pang, H., Kurjanczyk, L. A., and Penner, J. L. (1994) *Biochemistry* **33,** 241–249
14. Aspinall, G. O., Lynch, C. M., Pang, H., Shaver, R. T., and Moran, A. P. (1995) *Eur. J. Biochem.* **231,** 570–578
15. Nam Shin, J. E., Ackloo, S., Mainkar, A. S., Monteiro, M. A., Pang, H., Penner, J. L., and Aspinall, G. O. (1998) *Carbohydr. Res.* **305,** 223–232
16. Penner, J. L., Hennessy, J. N., and Congi, R. V. (1983) *Eur. J. Clin. Microbiol.* **2,** 378–383
17. Moran, A. P., and Penner, J. L. (1999) *J. Appl. Microbiol.* **86,** 361–377
18. Karlyshev, A. V., Linton, D., Gregson, N. A., Lastovica, A. J., and Wren, B. W. (2000) *Mol. Microbiol.* **35,** 529–541
19. Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V., Moule, S., Pallen, M. J., Penn, C. W., Quail, M. A.,

---

[3] St. Michael *et al.*, personal communication.

Rajandream, M.-A., Rutherford, K. M., van Vliet, A. H. M., Whitehead, S., and Barrell, B. G. (2000) *Nature* **403,** 665–669

20. Dingle, K. E., Colles, F. M., Wareing, D. R. A., Ure, R., Fox, A. J., Bolton, F. E., Bootsma, H. J., Willems, R. J. L., Urwin, R., and Maiden, M. C. J. (2001) *J. Clin. Microbiol.* **39,** 14–23

21. Suerbaum, S., Lohrengel, M., Sonnevend, A., Ruberg, F., and Kist, M. (2001) *J. Bacteriol.* **183,** 2553–2559

22. Linton, D., Gilbert, M., Hitchen, P. G., Dell, A., Morris, H. R., Wakarchuk, W. W., Gregson, N. A., and Wren, B. W. (2000) *Mol. Microbiol.* **37,** 501–514

23. Linton, D., Karlyshev, A. V., Hitchen, P. G., Morris, H. R., Dell, A., Gregson, N. A., and Wren, B. W. (2000) *Mol. Microbiol.* **35,** 1120–1134

24. Gilbert, M., Brisson, J.-R., Karwaski, M.-F., Michniewicz, J., Cunningham, A.-M., Wu, Y., Young, N. M., and Wakachuk, W. W. (2000) *J. Biol. Chem.* **275,** 3896–3906

25. Guerry, P., Ewing, C. P., Hickey, T. E., Prendergast, M. M., and Moran, A. P. (2000) *Infect. Immun.* **68,** 6656–6662

26. Wakarchuk, W. W., Martin, A., Jennings, M. P., Moxon, E. R., and Richards, J. C. (1996) *J. Biol. Chem.* **271,** 19166–19173

27. Gilbert, M., Watson, D. C., Cunningham, A.-M., Jennings, M. P., Young, N. M., and Wakarchuk, W. W. (1996) *J. Biol. Chem.* **271,** 28271–28276

28. Filiatrault, M. J., Gibson, B. W., Schilling, B., Sun, S., Munson, R. S., Jr., and Campagnari, A. A. (2000) *Infect. Immun.* **68,** 3352–3361

29. Prendergast, M. M., and Moran, A. P. (2000) *J. Endotoxin Res.* **6,** 341–359

30. van Belkum, A., van den Braak, N., Godschalk, P., Ang, W., Jacobs, B., Gilbert, M., Wakarchuk, W., Verbrugh, H., and Endtz, H. (2001) *Nat. Med.* **7,** 752–753

31. Jennings, M. P., Srikhanta, Y. N., Moxon, E. R., Kramer, M., Poolman, J. T., Kuipers, B., and van der Ley, P. (1999) *Microbiology* **145,** 3013–3021

32. Heinrichs, D. E., Yethon, J. A., and Whitfield, C. (1998) *Mol. Microbiol.* **30,** 221–232

33. Morona, J. K., Morona, R., and Paton, J. C. (1999) *J. Bacteriol.* **181,** 5355–5364

34. Jiang, S.-M., Wang, L., and Reeves, P. R. (2001) *Infect. Immun.* **69,** 1244–1255

35. Aspinall, G. O., Fujimoto, S., MacDonald, A. G., Pang, H., Kurjanczyk, L. A., and Penner, J. L. (1994) *Infect. Immun.* **62,** 2122–2125

36. Klena, J. D., Gray, S. A., and Konkel, M. E. (1998) *Gene* (*Amst.*) **222,** 177–185

**The Genetic Bases for the Variation in the Lipo-oligosaccharide of the Mucosal Pathogen, *Campylobacter jejuni* : BIOSYNTHESIS OF SIALYLATED GANGLIOSIDE MIMICS IN THE CORE OLIGOSACCHARIDE**

Michel Gilbert, Marie-France Karwaski, Stéphane Bernatchez, N. Martin Young, Eduardo Taboada, Joseph Michniewicz, Anna-Maria Cunningham and Warren W. Wakarchuk

Access the most updated version of this article at doi: 10.1074/jbc.M108452200

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 35 references, 11 of which can be accessed free at
http://www.jbc.org/content/277/1/327.full.html#ref-list-1