THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 19, pp. 14463–14475, May 11, 2007
Printed in the U.S.A.

Targeted Metabolomics Analysis of *Campylobacter coli* VC167 Reveals Legionaminic Acid Derivatives as Novel Flagellar Glycans*[S]

Received for publication, November 30, 2006, and in revised form, March 19, 2007 Published, JBC Papers in Press, March 19, 2007, DOI 10.1074/jbc.M611027200

David J. McNally[‡], Annie J. Aubry[‡], Joseph P. M. Hui[§], Nam H. Khieu[‡], Dennis Whitfield[‡], Cheryl P. Ewing[¶], Patricia Guerry[¶], Jean-Robert Brisson[‡], Susan M. Logan[‡1], and Evelyn C. Soo[§2]

From the [‡]National Research Council, Institute for Biological Sciences, Ottawa, Ontario K1A 0R6, Canada, the [§]National Research Council, Institute for Marine Biosciences, Halifax, Nova Scotia B3H 3Z1, Canada, and the [¶]Naval Medical Research Center, Silver Spring, Maryland 20910

Glycosylation of *Campylobacter* flagellin is required for the biogenesis of a functional flagella filament. Recently, we used a targeted metabolomics approach using mass spectrometry and NMR to identify changes in the metabolic profile of wild type and mutants in the flagellar glycosylation locus, characterize novel metabolites, and assign function to genes to define the pseudaminic acid biosynthetic pathway in *Campylobacter jejuni* 81–176 (McNally, D. J., Hui, J. P., Aubry, A. J., Mui, K. K., Guerry, P., Brisson, J. R., Logan, S. M., and Soo, E. C. (2006) *J. Biol. Chem.* 281, 18489–18498). In this study, we use a similar approach to further define the glycome and metabolomic complement of nucleotide-activated sugars in *Campylobacter coli* VC167. Herein we demonstrate that, in addition to CMP-pseudaminic acid, *C. coli* VC167 also produces two structurally distinct nucleotide-activated nonulosonate sugars that were observed as negative ions at *m*/*z* 637 and *m*/*z* 651 (CMP-315 and CMP-329). Hydrophilic interaction liquid chromatography-mass spectrometry yielded suitable amounts of the pure sugar nucleotides for NMR spectroscopy using a cold probe. Structural analysis in conjunction with molecular modeling identified the sugar moieties as acetamidino and *N*-methylacetimidoyl derivatives of legionaminic acid (Leg5Am7Ac and Leg5AmNMe7Ac). Targeted metabolomic analyses of isogenic mutants established a role for the *ptmA–F* genes and defined two new *ptm* genes in this locus as legionaminic acid biosynthetic enzymes. This is the first report of legionaminic acid in *Campylobacter* sp. and the first report of legionaminic acid derivatives as modifications on a protein.

Campylobacter sp. are among the most frequent cause of bacterial diarrhea worldwide and the leading cause of foodborne illness in North America (1, 2). Motility is critical to intestinal colonization by *Campylobacter* and is required for invasion of epithelial cells *in vitro*. *Campylobacter* flagella also function as secretory organelles in the absence of specialized type III secretion systems in this pathogen (3, 4). As such, flagella are recognized as a major virulence factor for *Campylobacter* sp. (5). The flagellar filament structural protein, FlaA, is the immunodominant protein recognized during infection and has been shown to be an immunoprotective antigen (6–9). Flagellins from numerous strains of *Campylobacter jejuni* and the related organism *Campylobacter coli* have been shown to be among the most heavily glycosylated prokaryotic proteins described (10–12), and the glycosyl modifications appear to be surface-exposed in the assembled filament and highly immunogenic (13, 14). In addition, it appears that unique forms of these glycosyl modifications contribute to the serospecificity of the flagellar filament (11) and the glycans are responsible for autoagglutination and microcolony formation (15).

All genes known to be involved in glycosylation of *Campylobacter* flagellins map near the *flaA* and *flaB* structural genes in a region that is one of the most hypervariable in the chromosome. The flagellin glycosylation locus can vary in size from ~25 kb in *C. jejuni* 81–176 to over 50 kb in *C. jejuni* strains NCTC11168 and RM1221 (16, 17). Our understanding of the nature of the glycans decorating flagellin has expanded considerably over the past few years. Mass spectrometry and NMR spectroscopy experiments showed definitively that the major sugars decorating flagellin from *C. jejuni* 81–176 were pseudaminic acid, 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid (Pse5Ac7Ac),[3] 5-acetamido-

* The work was supported by the National Research Council of Canada, NIAID, National Institutes of Health Grant RO1 AI43559, and the Military Infectious Disease Work Unit 6000.RAD1.DA3.A0308 (to P. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) EF141522.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S6 and Table S1.

[1] To whom correspondence may be addressed: Institute for Biological Sciences, National Research Council, Rm. 3037, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada. Tel.: 613-990-0839; Fax: 613-952-9092; E-mail: susan.logan@nrc-cnrc.gc.ca.

[2] To whom correspondence may be addressed: Institute for Marine Biosciences, National Research Council, 1411 Oxford St., Halifax, Nova Scotia B3H 3Z1, Canada. Tel.: 902-426-0780; Fax: 902-426-9413; E-mail: evelyn.soo@nrc-cnrc.gc.ca.

[3] The abbreviations used are: Pse5Ac7Ac, 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid (also known as pseudaminic acid); Pse5Ac7Am, 5-acetamido-7-acetamidino-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid; CE, capillary electrophoresis; CE-ESMS, capillary electrophoresis electrospray ionization-mass spectrometry; HILIC-MS, hydrophilic interaction liquid chromatography-mass spectrometry; HMBC, heteronuclear multiple bond correlation; HSQC, heteronuclear single quantum coherence; Leg5Ac7Ac, 5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid; Leg5Am7Ac, 5-acetamidino-7-acetamido-3,5,7,9,-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid; Leg5AmNMe7Ac, 5-*E*/*Z*-*N*-(*N*-methylacetimidoyl)-7-acetamidino-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid; Neu5Ac, *N*-acetylneuraminic acid; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TDP, thymidine diphos-

TABLE 1
***C. coli* VC167 mutants used in this study**

Strain no.	Cj number[a]	Gene name[b]	Marker	Reference
PG1244	Cj1319	unknown	*aph3*	This work
PG1245	Cj1320	unknown	*aph3*	This work
PG2151	Cj1324	*ptmG*	*cat*	This work
PG1541	Cj1325	*ptmH*	*cat*	This work
PG1273	Cj1327	*ptmC*	*aph3*	Logan *et al.* (11)
PG1214	Cj1328	*ptmD*	*aph3*	Logan *et al.* (11)
PG1540	Cj1329	*ptmE*	*cat*	Logan *et al.* (11)
PG1551	Cj1330	*ptmF*	*cat*	Logan *et al.* (11)
PG0917	Cj1331	*ptmB*	*aph3*	Guerry *et al.* (47)
PG0907	Cj1332	*ptmA*	*aph3*	Guerry *et al.* (47)
PG2657	Cj1121	*pglE*	*cat*	This work

[a] Refers to the homolog in the genome of NCTC11168 (16).
[b] Genes in the flagellin glycosylation locus of *C. coli* VC167 have been deposited at GenBank™ under accession number AY102621. The *pglE* gene of *C. coli* VC167 has been deposited in GenBank™ as accession number EF141522.

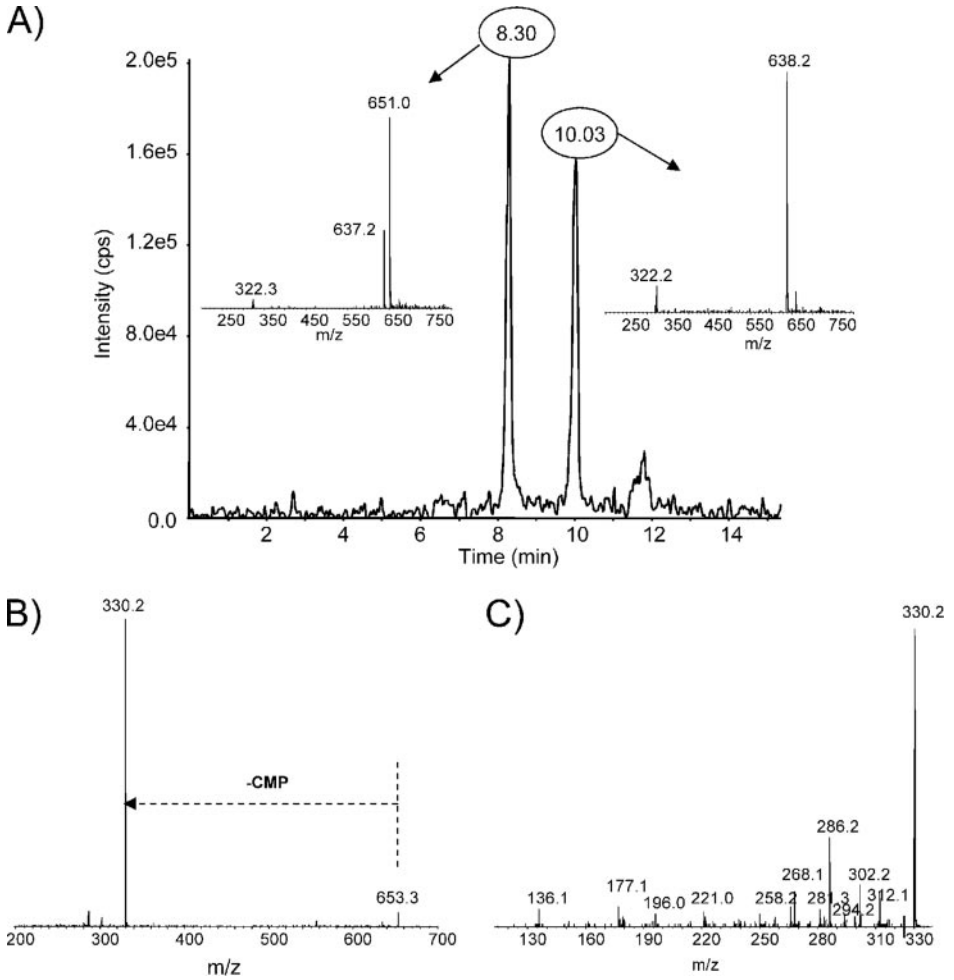

FIGURE 1. **Metabolomic analysis of *C. coli* VC167.** *A*, total ion chromatogram (*m/z* 100–1000) of CE-ESMS and precursor ion scans for fragment ions related to CMP (*m/z* 322). The *insets* show the corresponding mass spectra at 8.30 and 10.03 min, revealing the presence of two CMP-nonulosonic acids at *m/z* 637 and 638 and a novel CMP-sugar at *m/z* 651. *B*, MS/MS spectrum (positive mode) of the novel CMP-sugar observed as oxonium ions at *m/z* 653, showing fragment ions at *m/z* 330 that correspond to the carbohydrate moiety of the novel CMP-sugar (CMP-329). *C*, product ion spectrum of the carbohydrate moiety observed at *m/z* 330 obtained during MS^3 experiments (positive mode) on the novel CMP-sugar.

7-acetamidino-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid (Pse5Ac7Am), and related derivatives (10, 11, 18). Genetic analysis has identified the *pse* genes involved in Pse biosynthesis in *C. jejuni* 81–176 (10, 15, 18), and the enzymatic pathway for biosynthesis has been determined (19, 20). The *pse* pathway appears conserved among *Campylobacter* members. Genes in this pathway have also been characterized in *C. coli* VC167 (11), although the *pseA* gene, which is involved in synthesis of Pse in *C. jejuni* 81–176, is a pseudogene in this strain.

The structural characterization of flagellar modifications on *C. coli* VC167 has also been reported, and evidence was presented indicating that, although there were some groups conserved with *C. jejuni* 81–176, there were also differences in glycosyl modifications that could be distinguished serologically. In addition to Pse, *C. coli* VC167 appears to synthesize a structurally and serologically distinct acetamidino sugar having the same mass (315 Da) as the Pse5Ac7Am modification found on flagellin from *C. jejuni* 81–176 (11). Mutation of genes from the *ptm* locus in *C. coli* VC167 resulted in altered isoelectric focusing patterns of flagellin. Further, mass spectrometric analyses confirmed the loss of this 315-Da modification on the flagellin protein (11) thus implicating the *ptm* genes in the production of this flagellar glycan. Although the *ptm* locus is present in most strains of *C. jejuni* (with the exception of 81–176), the precise reason for the serological specificity of this unique 315-Da modification and the precise function of each of the *ptm* genes has yet to be determined.

The application of a novel CE-ESMS and precursor ion scanning method (21) in metabolomics studies of *Helicobacter pylori* (22) and *C. jejuni* 81–176 provided insight into the roles

phate; TOCSY, total correlation spectroscopy; UDP-QuiNAc4NAc, UDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose.

of genes from the flagellin glycosylation locus of these two bacterial pathogens and identified novel nucleotide-activated intermediates. McNally and coworkers (18) were also able to exploit this focused metabolomics approach and use *C. jejuni* 81–176 and the isogenic mutant strain *pseC* as a source for large-scale purifications of biosynthetic sugar-nucleotides relevant to the flagellin glycosylation process for precise structural analysis by NMR. CMP-Pse5Ac7Am was identified from the metabolome of wild-type *C. jejuni* 81–176. For the isogenic mutant *pseC*, UDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose (UDP-QuiNAc4NAc), a derivative of bacillosamine, was identified revealing an unexpected cross-talk between the *N*-linked glycosylation *pgl* pathway and the *O*-linked flagellar glycosylation *pse* pathway in *C. jejuni* 81–176 (18, 20).

The success of using targeted metabolomics strategies to elucidate unknown gene functions and identify novel biosynthetic substrates prompted us to employ this strategy to investigate the *ptm* flagellin glycosylation locus using *C. coli* VC167 as a model strain.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions—*C. coli* VC167 (23) has been described, and mutants of this strain are listed in Table 1. All *Campylobacter* strains were grown in Mueller-Hinton broth under microaerophilic conditions for 24 h at 37 °C. Media was supplemented with kanamycin (25 μg/ml) or chloramphenicol (10 μg/ml) when appropriate. For targeted metabolomic analysis of the parent strain *C. coli* VC167 and the isogenic mutants by CE-ESMS, cells from 500 ml of overnight culture were used. For purification of the intracellular sugar-nucleotide metabolites from *C. coli* VC167, cells from 10 liters of overnight culture of the parent strain were utilized.

Mutant Construction—Mutants in *C. coli* VC167 alleles of Cj1319 and Cj1320 were generated by insertion of an *aph3* cassette (24) into unique PstI and MluI sites within each gene, respectively. Mutation in *ptmG* and *ptmH* was done by transposition of a modified Tn5-transposon containing the *Campylobacter cat* gene as previously described (15). The insertion within *ptmG* occurred at bp 256 (in the 971-bp open reading frame) and that in *ptmH* at bp 520 of the 681-bp open reading frame. Mutation of the *pglE* gene was also done by Tn5 transposition. The insertion selected mapped to bp 27 within the *pglE* open reading frame. Plasmids from *Escherichia coli* were used to electroporate *C. coli* VC167 to either Km^R or Cm^R. Transformants were screened by PCR using primers that mapped outside the insertion point of the antibiotic gene to confirm that the DNA had undergone a double crossover.

Purification of Flagellin—Flagellin was purified as previously described (13).

Top Down MS Analysis of Flagellin—Flagellin was dialyzed in H_2O (0.2% formic acid) using a Centricon YM30 membrane filter. Flagellin was concentrated to 0.2 mg/ml and infused into a Waters Q-TOF mass spectrometer at a flow rate of 0.5 μl/min. Multiply protonated flagellin precursor ions were subjected to top down analysis according to Schirm *et al.* (28).

Preparation of Cell Lysates—Cell lysates were obtained as described in earlier studies (21, 22). Extraction of intracellular sugar-nucleotides from parent strain and isogenic mutants was achieved using Envi-Carb solid-phase extraction cartridges as described previously (18).

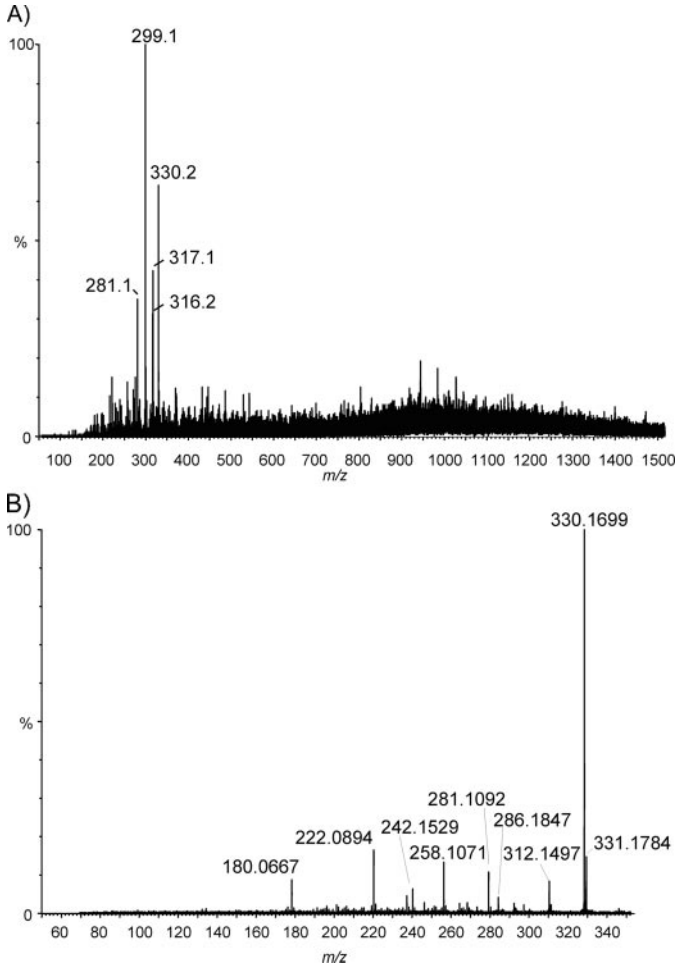

FIGURE 2. **Top down analysis of *C. coli* VC167 flagellin.** *A*, MS/MS spectrum of a multiply charged ion from intact flagellin acquired at 25-V collision energy using argon as the collision gas. Intense oxonium ions at *m/z* 330 corresponding to the 329-Da modification, at *m/z* 316 corresponding to the 315-Da modification, and at *m/z* 317, 299, and 281 corresponding to the 316-Da modification were observed in the spectrum. *B*, second generation fragment ion spectrum of the oxonium ion at *m/z* 330. Note that the MS/MS spectrum was acquired by increasing the RF Lens 1 from 50 to 125 forming fragment ions in the orifice/skimmer region to promote the formation of oxonium ions from the native flagellin.

Metabolomic Analysis by CE-ESMS—Cell lysates from parent strain *C. coli* VC167 and isogenic mutants *ptmA–F* were probed for intracellular sugar-nucleotides using a CE-ESMS and precursor ion scanning method as described earlier (21). The CE-MS instrumentation used in this study consisted of a CE system (Agilent Technologies, Santa Clara, CA) coupled to a 4000 QTRAP mass spectrometer equipped with a TurboV source via a coaxial sheath flow interface (AB/Sciex, Concord, Canada).

HILIC-MS Purifications of Sugar-Nucleotide Metabolites—The CMP-activated precursors were isolated from cell lysates of parent strain *C. coli* VC167 by HILIC-MS as described in earlier work (18). The liquid chromatography-MS instrumentation consisted of a 1100 Series LC system (Agilent Technologies) coupled to a 4000 QTRAP mass spectrometer (AB/Sciex). A TSKgel Amide80 column (4.6- × 250-mm inner diameter, Tosoh Bioscience, Montgomeryville, PA) was employed for the

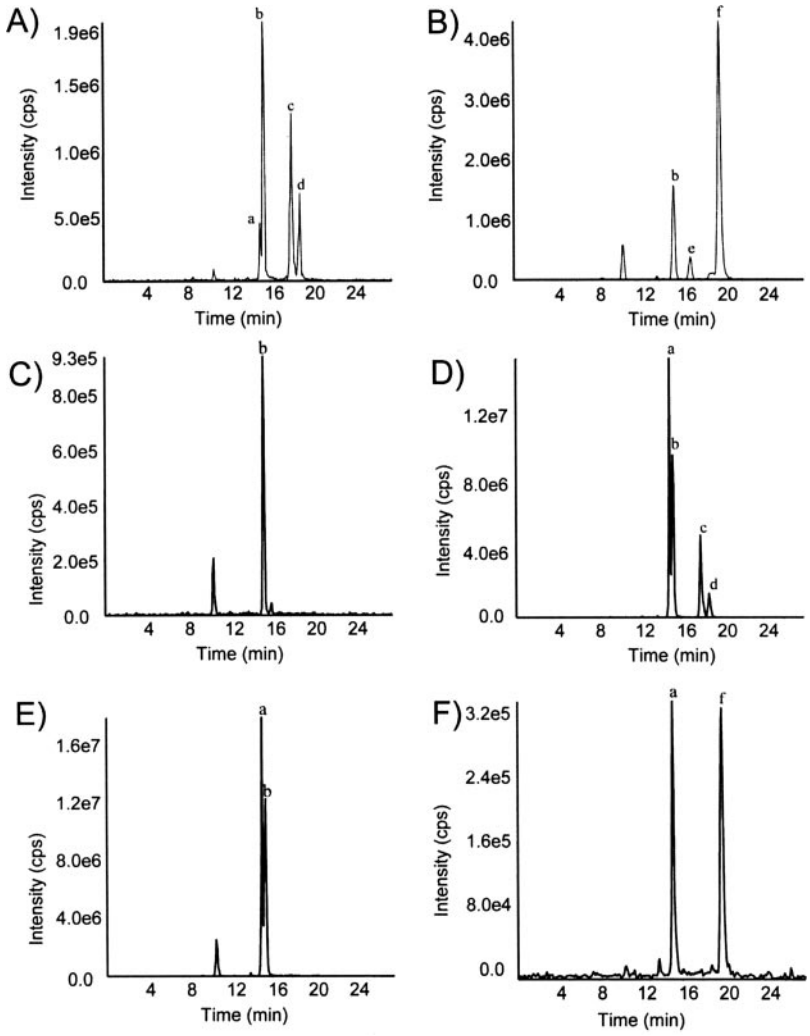

FIGURE 3. **HILIC-MS and precursor ion scanning for fragment ions related to CMP (*m/z* 322).** *A*, cell lysate from *C. coli* VC167; *B*, cell lysate from *C. jejuni* 81–176; *C*, isogenic mutant *ptm*C showing only the presence of *peak b*; *D*, *ptmC* complemented in *trans* (*ptmC:ptmC*) showing the restoration *peaks a*, *c*, and *d*; *E*, isogenic mutant of *ptmG* showing loss of *peaks c* and *d*; *F*, *ptmG* complemented with *pseA* (*ptmG:pseA*) showing the conversion of *peak b* to *peak f*. *Peak a*, CMP-Leg5Ac7Ac; *peak b*, CMP-Pse5Ac7Ac; *peak c*, CMP-Leg5AmNMe7Ac; *peak d*, CMP-Leg5Am7Ac; *peak e*, CMP-Neu5Ac; and *peak f*, CMP-Pse5Ac7Am.

HILIC separations, and the mobile phase was delivered to the column at a flow rate of 1.0 ml/min and split 2:8 v/v post-column via a stainless steel tee to the 4000 QTRAP mass spectrometer during fraction collection of the intracellular sugar-nucleotide intermediates. Selective detection of the intracellular CMP-linked sugars was achieved using precursor ion scanning for fragment ions related to CMP (*m/z* 322). An average of 300 fractions of each of the purified intermediates was pooled to achieve sufficient material for the NMR analysis.

NMR Spectroscopy of Sugar-Nucleotide Metabolites—Metabolites were lyophilized, resuspended in 200 μl of 99% D_2O (Cambridge Isotopes Laboratories Inc., Andover, MD), and analyzed by NMR spectroscopy. To observe exchangeable NH protons, metabolite samples were suspended in 95% H_2O (5% D_2O), and the pH was lowered through the addition of diluted DCl to reduce the rate of proton exchange (pH 3.7). To remove CMP from compound **II** forming Leg5Am7Ac (**III**), an aliquot of **II** was treated overnight with diluted DCl at 70 °C (pH 1.3). All samples were analyzed in 3-mm NMR tubes. Standard homo- and heteronuclear correlated two-dimensional ^1H NMR, ^{13}C HSQC, HMBC, COSY, TOCSY, and NOESY pulse sequences from Varian (Palo Alto, CA) were used for general assignments. Selective one-dimensional TOCSY experiments with a Z-filter and one-dimensional NOESY experiments were used for complete residue assignments and measurement of proton coupling constants ($J_{H,H}$) and NOEs (25–27). NMR experiments were performed with a Varian 600 MHz (^1H) spectrometer equipped with a Varian 5-mm Z-gradient triple resonance (^1H, ^{13}C, and ^{15}N) cryogenically cooled probe (cold probe) and with a Varian Inova 500-MHz (^1H) spectrometer with a Varian Z-gradient 3-mm triple resonance (^1H, ^{13}C, and ^{31}P) probe. NMR experiments were typically performed at 25 °C with suppression of the HOD resonance at 4.78 ppm. For proton and carbon experiments, the methyl resonance of acetone was used as an internal reference (δ_H 2.225 ppm and δ_C 31.07 ppm).

Molecular Dynamics Simulations for II and III—Molecular dynamics modeling was used to verify NOEs and to measure dihedral angles. Molecular models for II and III were constructed using the Biopolymer module of the Insight II Software package (Accelrys Inc., San Diego, CA). All subsequent calculations were performed using the computation module running on a Sybyl 7.0 environment (Tripos Inc., St. Louis, MO), and atoms were assigned Tripos potentials and Gasteiger-Huckel charges. To avoid unfavorable atomic contacts, the systems were subjected to a 500-step energy minimization using a BFGS method. Molecular dynamics simulations were then performed in vacuum at 400 K for 1000 ps following a 100-ps equilibration. A Verlet algorithm with a 0.5-fs time step group-based non-bond method with a cutoff distance of 10 Å and a distance-dependent dielectric value of 2 was used for simulations with trajectory frames being saved every 1 ps.

*Molecular Modeling of the N-Methylated Acetamidino Groups for **IV** and **V***—Molecular modeling was used to determine the lowest energy *E* and *Z* conformers for the *N*-methylated acetamidino groups in **IV** and **V**. Structures were optimized using the Austin Model 1 (AM1) Hamiltonian as implemented in Hyperchem 6.0 starting from idealized ring structures for cyclohexane. These optimized structures were then re-optimized using the Amsterdam Density Functional-

TABLE 2
Metabolomic analysis of *C. coli* VC167 flagellar glycosylation genes by HILIC-MS and precursor ion scanning for fragment ions related to CMP

Strain[a]	CMP-Pse5Ac7Ac (I), peak b	CMP-Pse5Ac7Am, peak f	CMP-Leg5Am7Ac (II), peak d	CMP-Leg5AmNMe7Ac (IV/V), peak c	CMP-Leg5Ac7Ac,[b] peak a
C. jejuni 81–176	+	+	−	−	−
C. coli VC167	+	−	+	+	+
ptmA	+	−	−	Trace	−
ptmB	+	−	−	−	−
ptmC	+	−	−	−	−
ptmD	+	−	−	−	−
ptmE	+	−	−	−	−
ptmF	+	−	−	Trace	−
ptmD (pRY111/*ptmD*)	+	−	+	+	++[c]
ptmC (pRY111/*ptmC*)	+	−	+	+	++[c]
Orf1 (Cj1319)	+	−	+	+	++[c]
Orf2 (Cj1320)	++[c]	−	+	+	+
ptmG	+	−	−	−	+
ptmH	+	−	+	−	+
pseB	−	−	+	+	+
pseB:ptmD	−	−	−	−	−
pglE	+	−	+	+	+
ptmG (pRY111/*pseA*)	−	++[c]	−	−	+

[a] Mutants are listed in Table 1. Complements of *ptm* mutants were described in Logan *et al.* (11).
[b] CMP-316 in *C. coli* VC167 metabolome.
[c] Increased amount of this metabolite relative to parent *C. coli* VC167.

Density Functional Theory-Quantum Mechanics (ADF-DFT-QM) program version 2005. Optimized structures were checked using frequency calculations (analytical second derivatives) for true convergence. Structures that exhibited negative frequencies were re-optimized using the coefficients of the negative modes to adjust the Cartesian coordinates. Optimization and frequency calculations were then repeated until all modes were positive. All structures were optimized as internal coordinates using the triple zeta plus basis set. Full solvation was considered using the ADF default continuum solvation model parameterized to water.

RESULTS

Metabolomic Analysis of C. coli VC167 by CE-ESMS—Using CE-ESMS and precursor ion scanning, cell lysates from wild-type *C. coli* VC167 were found to contain an intracellular pool of CMP-linked sugars at *m*/*z* 638 and 637 that were indicative of CMP-nonulosonic acids as observed before (18) (*insets*, Fig. 1*A*). However, in addition to these expected intermediates, the precursor ion-scanning experiments also revealed a novel intermediate as negative ions at *m*/*z* 651 that corresponds to a CMP-linked sugar of mass 329 Da. This mass is determined following subtraction of the mass of the CMP-carrier (*i.e.* *m*/*z* 651-322 (CMP + H_2O) = 329). To investigate this metabolite further, a series of tandem mass spectrometry experiments was performed. The product ion scan (negative mode) of the parent ions revealed a fragment ion at *m*/*z* 322 that is characteristic of the CMP moiety (data not shown). In the positive mode, oxonium ions corresponding to CMP-329 were observed at *m*/*z* 653, and tandem mass spectrometry experiments gave rise to a product ion corresponding to the 329 Da carbohydrate moiety at *m*/*z* 330 (Fig. 1*B*). To obtain structural information on this carbohydrate moiety, MS^3 experiments were performed on the novel CMP-sugar. The product ion spectrum (Fig. 1*C*) revealed losses of water that are typically observed with carbohydrates and an unusual fragment ion at *m*/*z* 258.1 that corresponded to the loss of 72.1 Da.

Top Down MS Analysis of C. coli VC167 Flagellin—Earlier analysis by LC-ESMS and MS/MS of tryptic digests of *C. coli* VC167 flagellin (bottom up approach) revealed the presence of both 315- and 316-Da glycans on the flagellin protein (11). These modifications correspond to the mass of the respective glycan moieties of the CMP-nonulosonic acids (*m*/*z* 637 and 638), which we found in the metabolome of *C. coli* VC167. In contrast, the 329-Da carbohydrate moiety was not observed as a modification of the flagellin protein in these initial bottom up studies. It is known that there are limitations to the use of a bottom up approach in the identification and characterization of post-translational modifications. For example, low stoichiometric abundance of a particular modification and intrinsic properties leading to poor ionization efficiency or instability during digestion and sample preparation may prevent the identification of novel glycan moieties. MS analysis of an intact protein (top down approach) is now often used in conjunction with the bottom up approach to provide a more complete analysis of a post-translationally modified protein. When this type of analysis was performed in the current study, the presence of a glycan moiety of mass 329 was clearly demonstrated on the flagellin protein. As can be seen in Fig. 2*A*, in addition to the expected modifications giving oxonium ions at 315 and 316 Da (*m*/*z* 316 and 317), an oxonium ion of *m*/*z* 330 was observed. This oxonium ion corresponds to the 329-Da glycosyl moiety that had been found to be CMP-activated (*m*/*z* 651) in the metabolome during CE-ESMS and precursor ion-scanning experiments. In addition, fragment ions at *m*/*z* 299 and 281, which result from neutral losses of water from *m*/*z* 317, were also observed in the mass spectrum. The MS/MS spectrum of the oxonium ion at *m*/*z* 330 from the flagellin protein (Fig. 2*B*) provided additional confirmation that this glycosyl moiety corresponds to the glycosyl moiety of the CMP-329 metabolite (Fig. 1*C*). HILIC-MS and NMR experiments were undertaken to fully characterize this unknown carbohydrate, suspected to be related to a nonulosonic acid.

HILIC-MS Purification of C. coli VC167 CMP-activated Metabolites—The unique selectivity of the HILIC stationary phase was used to resolve the metabolome of *C. coli* VC167.

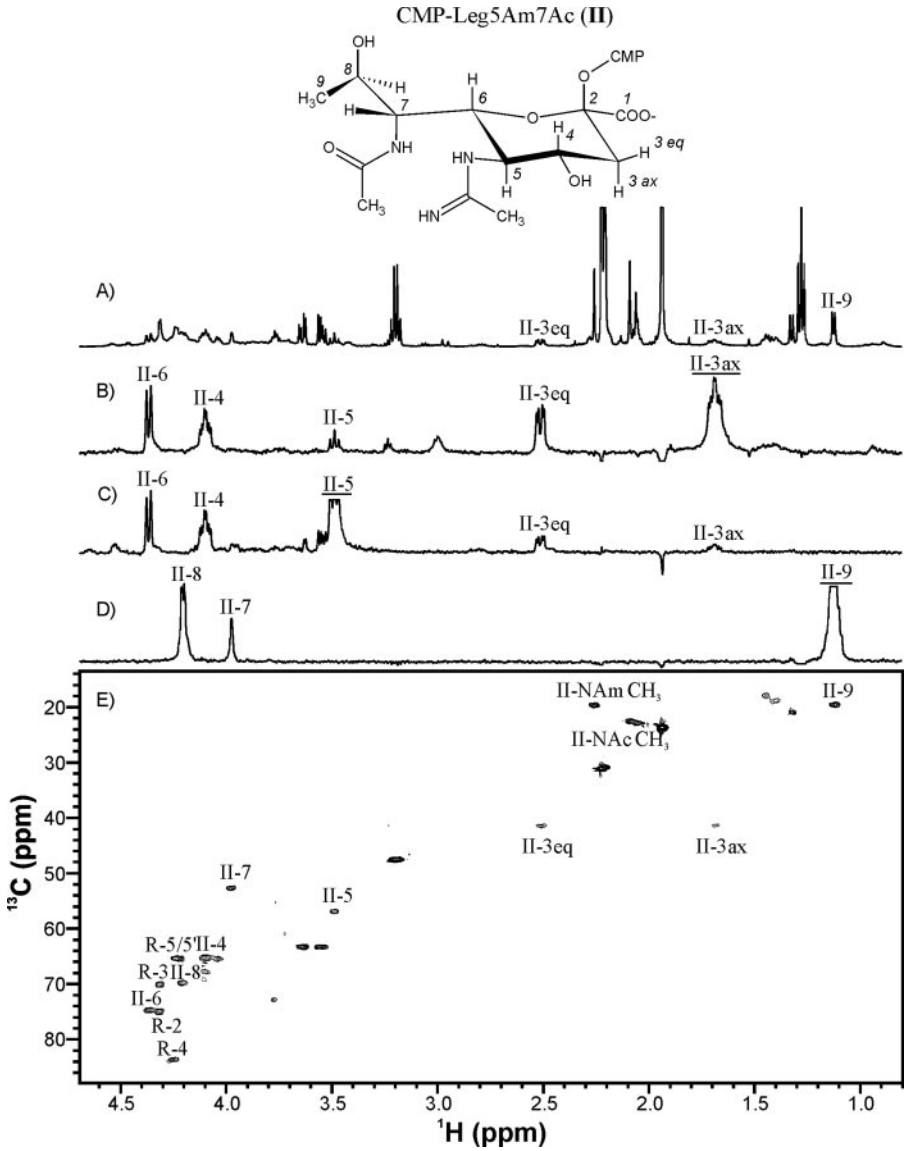

FIGURE 4. **NMR spectroscopy of CMP-Leg5Am7Ac (II).** *A*, ^{1}H NMR spectrum (1024 transients); *B–D*, one-dimensional TOCSY (80 ms) of H3ax, H5, and H9 resonances (*underlined*); *E*, ^{13}C HSQC spectrum (384 transients, 128 increments, $^1J_{C,H}$ = 140 Hz, and 7.5 h). NMR experiments were performed in D_2O (pD 7.2, 25 °C) at 500 MHz (^{1}H) for *A–D* and at 600 MHz (^{1}H) with a cold probe for *E*. For *panel E*, *R* represents ribose.

The two peaks containing CMP-activated sugars observed by CE-ESMS (Fig. 1*A*) were resolved into four peaks by HILIC-MS (Fig. 3*A*). This method of analysis also revealed differences between the metabolome of *C. coli* VC167 and *C. jejuni* 81–176 as presented in Fig. 3 and Table 2. For *C. coli* VC167, four peaks (*a*, *b*, *c*, and *d*) were observed (Fig. 3*A*). For *C. jejuni* 81–176, peaks *b*, *e*, and *f* were observed (Fig. 3*B*). From previous studies for *C. jejuni* 81–176, peaks *b*, *e*, and *f* were identified as CMP-Pse5Ac7Ac, CMP-Neu5Ac, and CMP-Pse5Ac7Am, respectively (18). The mass spectrum of peak *d* revealed a *m*/*z* 637 ion, and the HILIC-MS data clearly revealed this as a novel metabolite (CMP-315) with a unique retention time (**II**). A small amount of a second CMP-316 metabolite (peak *a*) at *m*/*z* 638 was also observed in the HILIC separation of the *C. coli* VC167 metabolome that had not been resolved by CE-ESMS. Tandem mass spectrometry experiments indicated that this metabolite was related to the CMP-315 metabolite from the *C. coli* VC167 metabolome. Separate fractions of CMP-Pse5Ac7Ac (**I**, peak *b*), CMP-315 (**II**, peak *d*), and CMP-329 (**IV**/**V**, peak *c*) from *C. coli* VC167 were pooled, dried on a SpeedVac concentrator, and stored at −20 °C for later analysis by NMR.

*Structural Elucidation of **I** and **II** by NMR Spectroscopy*—Based on CE-ESMS experiments that indicated CMP-linked nonulosonic acid metabolites, two-dimensional COSY and TOCSY (90 ms) experiments were used to locate resonances that are characteristic of these CMP-linked nonulosonic acid metabolites such as H3ax, H3eq, and H9 signals (not shown). A combination of one- and two-dimensional homonuclear (^{1}H-^{1}H) and heteronuclear (^{13}C-^{1}H) experiments were then used to elucidate their complete structures. Proton and carbon chemical shifts as well as couplings constants for **I** were found to be in excellent agreement with those obtained for an authentic CMP-Pse5Ac7Ac standard that was enzymatically synthesized in our laboratory (20). This confirmed the identity of **I** as CMP-Pse5Ac7Ac (supplemental Fig. S1).

The proton spectrum for **II** revealed broad H3ax and H3eq signals and sharp signals at δ_H 2.09 and 2.26 ppm that indicated the presence of an acetamido and acetamidino group, respectively (Fig. 4*A* and Table 3). One-dimensional TOCSY experiments were then used to assign chemical shifts and to measure coupling constants. These experiments are particularly useful for assigning overlapping resonances originating from sugar ring protons and for resolving mixtures of compounds (25–27). One-dimensional TOCSY of H3ax (Fig. 4*B*) and H5 (Fig. 4*C*) revealed spin-coupled signals corresponding to H4, H5, and H6. The large $J_{4,5}$ (10.0 Hz), $J_{5,6}$ (10.0 Hz), and a small $J_{6,7}$ coupling of 1.6 Hz were indicative of legionaminic acid (29–31). The large $^4J_{P,H3ax}$ coupling (4.8 Hz) indicated the β anomer (32). In contrast to reports for the monosaccharide of legionaminic acid, one-dimensional TOCSY of H9 revealed a small $J_{7,8}$ coupling of 2.8 Hz (Fig. 4*D*). Upon removal of the CMP group by acid hydrolysis to form **III**, $J_{7,8}$ was determined to be 7.6 Hz (supplemental Fig. S2), in good agreement with the value reported for legionaminic acid (33). To establish the location of the acetamidino group, **II** was suspended in 95% H_2O (5% D_2O), and the pH was lowered through the addition of diluted DCl (pH 3.7) (18). The H5-5NH COSY correlation indicated the location of the acetamidino group at

TABLE 3

NMR data for CMP-Leg5Am7Ac (II) and Leg5Am7Ac (III)

CMP assignments are not shown. $5NAm^{CH3}$ and $5NAm^{C=N}$ represent the *C*-methyl group and nonprotonated carbon of the 5-acetimidino group, respectively. $7NAc^{CH3}$ and $7NAc^{C=O}$ represent the *C*-methyl and nonprotonated carbon of the 7-acetamido group, respectively. Chemical shifts for exchangeable NH protons were determined in 95% H_2O (5% D_2O) at pH 3.7 for **II** and at pH 1.3 for **III**. Carbon and proton chemical shifts were referenced to an internal acetone standard (δ_H 2.225 ppm and δ_C 31.07 ppm). Error for δ_H is ±0.02 ppm, for δ_C is ±0.2 ppm, and for $J_{H,H}$ is ±0.2 Hz.

Compound	Proton (^{1}H)	δ_H	Carbon (^{13}C)	δ_C	$J_{H,H}$	Hz
		ppm		*ppm*		
II			C1	ND[a]	$J_{3ax,3eq}$	13.3
			C2	100.6	$J_{3ax,4}$	12.2
	H3ax	1.69	C3	41.3	$J_{P,3ax}$	4.8
	H3eq	2.52			$J_{3eq,4}$	4.7
	H4	4.10	C4	67.8	$J_{4,5}$	10.0
	H5	3.49	C5	56.9	$J_{5,6}$	10.0
	H6	4.37	C6	74.6	$J_{6,7}$	1.6
	H7	3.98	C7	52.6	$J_{7,8}$	2.8
	H8	4.21	C8	69.7	$J_{8,9}$	6.4
	H9	1.12	C9	19.6		
	5NH	9.28			$J_{NH,5}$	9.9
	$5NAm^{CH3}$	2.26		19.6		
			$5NAm^{C=N}$	167.7		
	$5NAm^{NH2}$	8.29/8.73				
	7NH	8.81			$J_{NH,7}$	10.4
	$7NAc^{CH3}$	2.09		22.5		
			$7NAc^{C=O}$	175.2		
III			C1	ND		
			C2	ND	$J_{3ax,3eq}$	13.2
	H3ax	1.90	C3	39.9	$J_{3ax,4}$	11.9
	H3eq	2.32			$J_{3eq,4}$	5.0
	H4	4.07	C4	68.3	$J_{4,5}$	10.0
	H5	3.51	C5	57.4	$J_{5,6}$	10.0
	H6	4.24	C6	71.6	$J_{6,7}$	1.9
	H7	3.94	C7	54.8	$J_{7,8}$	7.6
	H8	3.92	C8	68.8	$J_{8,9}$	5.9
	H9	1.20	C9	19.9		
	5NH	9.21			$J_{NH,5}$	9.8
	$5NAm^{CH3}$	2.26		20.1		
			$5NAm^{C=N}$	168.0		
	$5NAm^{NH2}$	8.29/8.72				
	7NH	8.50			$J_{NH,7}$	9.2
	$7NAc^{CH3}$	2.05		23.1		
			$7NAc^{C=O}$	175.5		

[a] ND, not determined.

H5 (18, 34). Further, compared with the monosaccharide of legionaminic acid (33), C5 was found to resonate downfield by 3.3 ppm and is in good agreement with the effects reported for substitution with an acetamidino group (30, 35).

The absolute configuration of **II** and **III** was determined by comparison of their NMR data with those of known legionaminic acids (31, 33) and molecular modeling. Comparison of chemical shift data for **III** indicated that the absolute configuration was either D-*glycero*-D-*galacto*-legionaminic acid or its 8-epimer, L-*glycero*-D-*galacto*-legionaminic acid. Based on a strong H9-H7 NOE observed for **II** and **III** (supplemental Table S1 and Fig. S3), the absolute configuration was determined to be D-*glycero*-D-*galacto*. Molecular modeling was used to measure interproton distances for both absolute configurations and confirmed a strong H9-H7 and H8-H6 NOE and a weaker H9-H6 NOE for D-*glycero*-D-*galacto*. Interestingly, substantial differences for $J_{7,8}$ values and C6, C7, and C8 carbon chemical shifts were observed for **II** and **III** (Table 3). Simulations of molecular dynamics trajectories (400 K) for **II** and **III** showed that the bulky CMP group exerts a long range effect on the H7-C7-C8-H8 torsional angle. In contrast, the H6-C6-C7-H7 torsional angle was found to be much less affected by the presence of the nucleotide (supplemental Fig. S4). These simulations thus support that CMP influences the conformation about the C7-C8 bond and explain the differences observed for $J_{7,8}$ and C6, C7, and C8 chemical shifts in **II** and **III**. Based on the combined results, **II** was identified as CMP-5-acetamidino-7-acetamido-3,5,7,9,-tetradeoxy-D-*glycero*-β-D-*galacto*-nonulosonic (CMP-Leg5Am7Ac) and **III** as 5-acetamidino-7-acetamido-3,5,7,9,-tetradeoxy-D-*glycero*-β-D-*galacto*-nonulosonic acid (Leg5Am7Ac).

*Structural Elucidation of **IV** and **V** by NMR Spectroscopy*—Two-dimensional COSY and TOCSY (90 ms) experiments (not shown) of the CMP-329 sample (*peak c*) revealed two distinct sets of spin-correlated resonances in equal proportion corresponding to compounds **IV** and **V**. Proton chemical shifts for **IV** and **V** were highly similar to **II** and thus indicated that these compounds were structurally similar derivatives of legionaminic acid (Tables 3 and 4). COSY correlations observed between the NAm^{CH3} signals and broad CH_3 singlets near 3 ppm indicated that the acetamidino groups were *N*-methylated. MS^3 experiments of the CMP-329 sample (see above) supported that acetamidino groups were *N*-methylated, because fragment ions at *m*/*z* 258.1 were observed corresponding to the loss of a 72.1-Da moiety (Fig. 1*C*). One-dimensional TOCSY of **IV**/**V** H3ax signals revealed overlapping resonances for **IV**/**V** H4 and H5, and distinct H6 resonances (Fig. 5*B*). Overlapped signals were assigned through one-dimensional TOCSY of **IV** H6 that showed one clear set of resonances for **IV** H3ax, H3eq, H4, H5, and H7 (Fig. 5*C*). A one-dimensional TOCSY experiment of **IV**/**V** H9 signals indicated identical chemical shifts for H8 of both compounds (Fig. 5*D*), whereas a

TABLE 4

NMR data for *E*(IV) and *Z*(V) forms of CMP-Leg5AmNMe7Ac

CMP assignments are not shown. $5NAm^{NMe}$, $5NAm^{CH3}$, and $5NAm^{C=N}$ represent the *N*-methyl group, *C*-methyl group, and nonprotonated carbon of the 5-acetimidoyl group, respectively. $7NAc^{CH3}$ and $7NAc^{C=O}$ represent the *C*-methyl and nonprotonated carbon of the 7-acetamido group, respectively. Chemical shifts and coupling constants for exchangeable NH protons were determined in 95% H_2O (5% D_2O) at pH 3.7. Carbon and proton chemical shifts were referenced to an internal acetone standard (δ_H 2.225 ppm and δ_C 31.07 ppm). Error for δ_H is ± 0.02 ppm, for δ_C is ± 0.2 ppm, and for $J_{H,H}$ is ± 0.2 Hz.

Compound	Proton (^{1}H)	δ_H	Carbon (^{13}C)	δ_C	$J_{H,H}$	Hz
		ppm		*ppm*		
IV			C1	ND[a]	$J_{3ax,3eq}$	13.3
			C2	100.6	$J_{3ax,4}$	12.2
	H3ax	1.71	C3	41.3	$J_{P,3ax}$	5.2
	H3eq	2.55			$J_{3eq,4}$	4.7
	H4	4.14	C4	67.9	$J_{4,5}$	9.9
	H5	3.50	C5	58.4	$J_{5,6}$	10.1
	H6	4.54	C6	74.4	$J_{6,7}$	1.3
	H7	3.99	C7	52.6	$J_{7,8}$	2.8
	H8	4.20	C8	69.8	$J_{8,9}$	6.4
	H9	1.13	C9	19.5		
	5NH	8.19			$J_{NH,5}$	10.1
	$5NAm^{CH3}$	2.28		17.9		
			$5NAm^{C=N}$	167.2		
	$5NAm^{NH}$	8.50			$J_{NH,CH3}$	4.3
	$5NAm^{NMe}$	2.98		30.4		
	7NH	8.93			$J_{NH,7}$	10.2
	$7NAc^{CH3}$	2.13		22.5		
			$7NAc^{C=O}$	175.1		
V			C1	ND	$J_{3ax,3eq}$	13.5
			C2	100.6	$J_{3ax,4}$	12.0
	H3ax	1.69	C3	41.3	$J_{P,3ax}$	5.2
	H3eq	2.52			$J_{3eq,4}$	4.8
	H4	4.10	C4	67.9	$J_{4,5}$	9.8
	H5	3.47	C5	56.1	$J_{5,6}$	9.9
	H6	4.36	C6	74.8	$J_{6,7}$	1.5
	H7	3.98	C7	52.6	$J_{7,8}$	2.9
	H8	4.20	C8	69.8	$J_{8,9}$	6.4
	H9	1.13	C9	19.5		
	5NH	9.08			$J_{NH,5}$	10.1
	$5NAm^{CH3}$	2.18		18.4		
			$5NAm^{C=N}$	166.7		
	$5NAm^{NH}$	9.10			$J_{NH,CH3}$	4.1
	$5NAm^{NMe}$	2.95		29.4		
	7NH	8.78			$J_{NH,7}$	10.1
	$7NAc^{CH3}$	2.09		22.5		
			$7NAc^{C=O}$	175.3		

[a] ND, not determined.

^{13}C HSQC experiment was used to assign carbon resonances (Fig. 5*E*). The anomeric and absolute configurations of **IV** and **V** were determined to be the same as **II** based on similar proton chemical shifts, carbon chemical shifts, coupling constants, and NOEs (supplemental Table S1 and Fig. S3).

The ^{13}C chemical shifts for **IV** and **V** were very similar, except for C5, indicating that they differed only with regards to their *N*-methylated groups. NMR experiments performed in 95% H_2O and at lower pH (pH 3.7) revealed exchangeable NH resonances and confirmed the location of the *N*-methylated acetamidino group at C5 (data not shown). By comparing carbon chemical shifts to those reported for the *N*-methylated acetamidino structures in *Legionella pneumophila* (36), the $5NAm^{NMe}$ group was determined to be located on the terminal nitrogen atom (N^1) (Fig. 5). A strong $5NAm^{NMe}$/$5NAm^{C=N}$ HMBC correlation, the lack of the $5NAm^{NMe}$/C5 HMBC correlation, protonation of the N^2 atom ($5NAm^{NH}$), and coupling between $5NAm^{NH}$ and $5NAm^{NMe}$ ($J_{NH,CH3}$) confirmed that the *N*-methyl group was located at N^1. A two-dimensional NOESY experiment (800 ms) showed a strong $5NAm^{CH3}$-$5NAm^{NMe}$ NOE for **IV** and a much weaker one for **V** (supplemental Table S1). Based on these NOEs and a study that showed *N*-methylated acetamidino groups exist in solution as *Z* (*cis*) and *E* (*trans*) stereoisomers (36), the $5NAm^{NMe}$ group in **IV** was concluded to have an *E* orientation, whereas that in **V** was concluded to have a *Z* orientation (Fig. 5).

Since there is only limited information available for these *N*-methylated acetamidino groups (37, 38), molecular modeling was performed to corroborate the NMR results. Due to the partial double-bond character of the *N*-methylated acetamidino groups, two *E* conformations can be drawn for **IV** (*E1* and *E2*, Fig. 6, *A* and *B*), and two *Z* conformations can be drawn for **V** (*Z1* and *Z2*, Fig. 6, *C* and *D*). Furthermore, N^2 (5N) in both compounds can have one proton (monoprotonated) or two protons (diprotonated). Molecular modeling was first used to determine if N^2 is mono- or diprotonated and then to determine the lowest energy conformer for **IV** and **V**. By modeling idealized cyclohexane rings with protonated amidine (amidinium) groups (supplemental Fig. S5), it was established that N^2 is most likely monoprotonated and that this proton (also referred to as the 5NH proton) is *trans* with respect to the H5 ring proton. This latter finding is supported by the large $J_{NH,5}$ coupling measured for the 5NH proton (Table 4). Comparison of the relative energies for the molecular models revealed that **IV** most likely adopts the *E* conformer shown in Fig. 6*A* (4 kJ·mol^{-1}), whereas **V** adopts the *Z* conformer shown in Fig. 6*C* (0 kJ·mol^{-1}). This finding is supported by a strong NOE observed between the NAm^{CH3} and H5 in **V** and the absence of

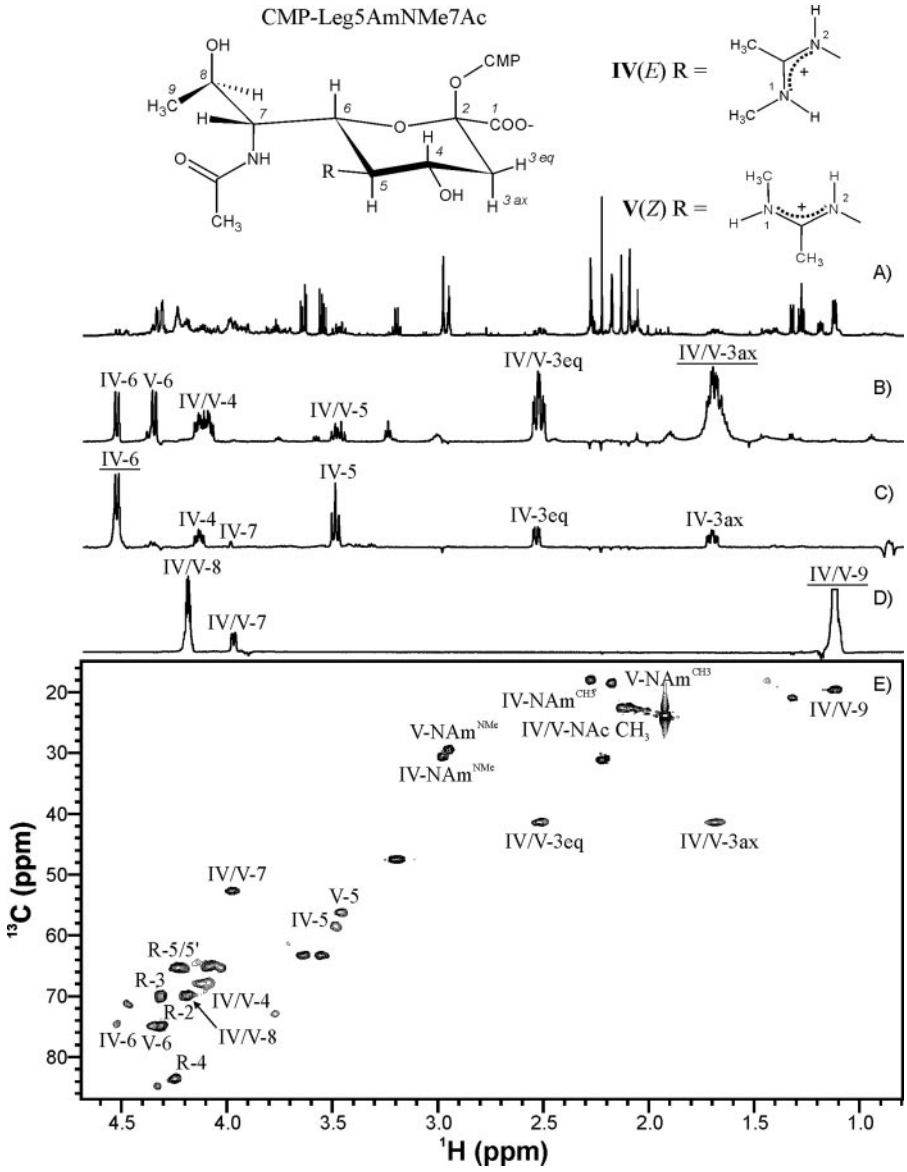

FIGURE 5. **NMR spectroscopy of *E* (IV) and *Z* (V) forms of CMP-Leg5AmNMe7Ac.** *A*, ^{1}H NMR spectrum (1024 transients); *B–D*, one-dimensional TOCSY (80 ms) of **IV**/**V** H3ax, **IV** H6, and **IV**/**V** H9 resonances (*underlined*); *E*, ^{13}C HSQC spectrum (384 transients, 128 increments, $^{1}J_{C,H}$ = 140 Hz, and 7.5 h). NMR experiments were performed in D_2O (pD 7.2, 25 °C) at 600 MHz (^{1}H) with a cold probe. For *panel E*, *R* represents ribose.

this NOE in **IV** (supplemental Table S1). Based on the small energy difference (4 kJ·mol^{-1}), it is expected that conversion between the *E* and *Z* forms is a frequent occurrence, in accord with equimolar amounts of **IV** and **V** in solution (Fig. 5). Interestingly, molecular modeling showed that the *N*-methylated acetamidino groups are planar and would be prominent structural features on flagellin protein. Collectively, these NMR and molecular modeling results established the identity of **IV** as CMP-Leg5- *E*-(*N*-methylacetimidoyl)7Ac, and **V** as CMP-Leg5-*Z*-(*N*-methylacetimidoyl)7Ac. Compounds **IV** and **V** will forthwith be referred to as CMP-Leg5AmNMe7Ac.

Metabolomic Analysis of C. coli VC167 Mutants—It had been previously shown that the region of the chromosome between *pseA* and the flagellin structural genes in *C. coli* VC167 contains the *ptm* genes. Mutation of these genes did not effect motility but resulted in flagellin with altered isoelectric focusing patterns, and mass spectrometry analysis had demonstrated that flagellin was no longer modified with the serologically distinct form of the 315-Da sugar, which had been shown in *C. jejuni* to be Pse5Ac7Am (10, 11, 15). Although none of the *ptm* genes are present in the genome of *C. jejuni* 81–176, homologs are found in a significant number of other *Campylobacter* species, including the genome sequenced strains *C. jejuni* NCTC 11168 and *C. jejuni* RM1221, as well as *C. coli* RM2228 and *C. jejuni* subsp *doylei* 269.97 (16, 17).

We have now confirmed that this serologically distinct 315-Da sugar made by *C. coli* VC167 is Leg5Am7Ac. In addition, we have demonstrated, by top down analysis, that the flagellin is also modified with a related glycan, Leg5AmNMe7Ac. By using HILIC-MS with a precursor ion-scanning approach we next investigated the role of the *ptm* locus in the biosynthesis of these distinct glycan moieties. Isogenic mutants *ptmA–F*, as well as isogenic mutants of *C. coli* VC167 in the other four open reading frames from this region (homologs of Cj1319, Cj1320, Cj1324, and Cj1325; see Table 1), were screened individually for intracellular sugar-nucleotides relating to the biosynthesis of Pse and Leg sugars (Table 2) to determine the role of each of these genes in the respective biosynthetic pathways of Leg5Am7Ac and Leg5AmNMe7Ac.

The insertional inactivation of *ptmC* inhibited the biosynthesis of Leg sugars as indicated by the disappearance of peaks *a*, *c*, and *d* in Fig. 3*C*. Absence of CMP-Leg5Am7Ac (peak *d*) is consistent with earlier observations where the flagellin of the isogenic mutant *ptmC* was shown to produce flagellin that was no longer glycosylated with a 315-Da glycan (11). Metabolomic analysis of the remaining *ptm* genes showed that they had metabolic profiles similar to that obtained for *ptmC* thereby confirming their role in the Leg pathway (Table 2). In contrast to the *C. coli* VC167 parent strain, only trace amounts of CMP-Leg5AmNMe7Ac was observed in the *ptmA* and *ptmF* mutants. Complementation in *trans* of *ptmD* (data not shown) and *ptmC* (Fig. 3*D*) restored the ability to synthesize Leg sugars as seen by the presence of peaks *a*, *c*, and *d*. Of note, these strains also produced an increased level of CMP-Leg5Ac7Ac (peak *a*), which had been observed in *C. coli* VC167 at trace levels (Fig.

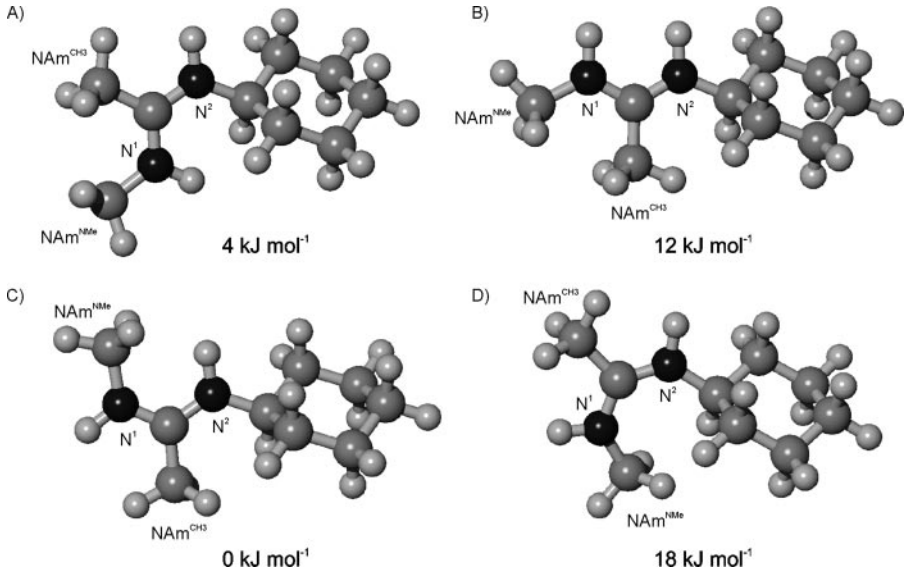

FIGURE 6. **Molecular modeling of the *E* and *Z* conformations of *N*-methylated acetamidino groups attached to idealized cyclohexane rings.** *A* and *B* show the lowest energy *E* conformations that are representative of the CMP-Leg5AmNMe7Ac structure **IV**, whereas *C* and *D* show the lowest energy *Z* conformations representative of the CMP-Leg5AmNMe7Ac structure **V**.

3*A*) and was most probably due to complementation in *trans* on plasmid pRY111 (39).

In addition to the previously described *ptm* genes, insertional inactivation of the homolog of Cj1324 also revealed a loss in the ability to synthesize both CMP-Leg5Am7Ac (**II**) and CMP-Leg5AmNMe7Ac (**IV**/**V**) providing the first evidence for a functional role of this gene in the Leg pathway. This gene encodes a protein with high homology to PseA from *C. jejuni* 81–176 that has been shown to be involved in the biosynthesis of CMP-Pse5Ac7Am from CMP-Pse5Ac7Ac in *C. jejuni* 81–176 (10, 18). Based on these data, this gene has been annotated as *ptmG* (see Table 1). As observed in Fig. 3*E*, the level of CMP-Leg5Ac7Ac (peak *a*) was increased in this mutant. Thus, in a manner analogous to PseA, it appears that the enzyme encoded by *ptmG* is critical for the biosynthesis of CMP-Leg5AmNMe7Ac and CMP-Leg5Am7Ac from CMP-Leg5Ac7Ac.

The specificity of the enzymatic products of *ptmG* and *pseA*, in the biosynthesis of CMP-Leg5Am7Ac and CMP-Pse5Ac7Am, respectively, was demonstrated in this current metabolomic analysis. Complementation of the *ptmG* mutant of *C. coli* VC167 with *pseA* from *C. jejuni* 81–176 (Fig. 3*F* and Table 2) resulted in the conversion of all CMP-Pse5Ac7Ac (peak *b*) in the metabolome to CMP-Pse5Ac7Am (peak *f*). There was no evidence for conversion of CMP-Leg5Ac7Ac (peak *a*) to CMP-Leg5Am7Ac (peak *d*, Fig. 3*F*). In *ptmA–F* mutants where *ptmG* was still functional, but cells are unable to synthesize any CMP-Leg sugar nucleotides, no conversion of CMP-Pse5Ac7Ac to CMP-Pse5Ac7Am was observed. Analysis of the metabolome of the isogenic mutant in the *C. coli* VC167 homolog of Cj1325, revealed only the absence of CMP-Leg5AmNMe7Ac, whereas production of CMP-Pse5Ac7Ac, CMP-Leg5Am7Ac, and CMP-Leg5Ac7Ac was unaffected. Based on these data, this gene has been annotated as *ptmH* (see Table 1).

Mutation of the homologs of Cj1319 and Cj1320 resulted in no change in the composition of the CMP-metabolite pool. However, the mutant in the Cj1319 homolog appeared to have a notable increase of CMP-Leg5Ac7Ac that had not been observed in any other of the mutants analyzed. As expected, mutation of *pseB*, the first enzyme of the *pse* pathway, prevented the production of CMP-Pse5Ac7Ac, whereas a double mutant in the *pseB* and *ptmD* genes resulted in an inability to produce any CMP-activated monosaccharides (Table 2). As previously reported, this mutant is non-motile (40).

Accumulation of Nucleotide-activated Intermediates—In earlier studies, metabolomic screening of *C. jejuni* 81–176 isogenic mutants in Pse biosynthetic genes revealed an accumulation of UDP-activated intermediates. *C. coli* VC167 mutants that had been affected in their ability to produce CMP-Leg nucleotides were screened for accumulation of UDP, ADP, TDP, and GDP activated intermediates. No accumulation of alternate nucleotide-activated metabolites was observed in any of the *C. coli* VC167 mutant strains (data not shown).

Role of pgl Biosynthetic Pathway Enzymes in the Biosynthesis of Legionaminic Acid Sugars—Because both pseudaminic and legionaminic acids are nonulosonate sugars, it seems reasonable that the Leg biosynthetic pathway may resemble the *pse* pathway (20). One possible biosynthetic route for the production of legionaminic acid could occur through the *pgl* pathway, which is responsible for the *N*-linked glycosylation of a number of proteins in *Campylobacter* cells (41). UDP-QuiNAc4NAc is synthesized by *pglE*, *-F*, and *-D* in this pathway (19, 42) (supplemental Fig. S6). 2,4-Diacetamido-2,4,6-trideoxy-α-D-mannose could be generated from UDP-QuiNAc4NAc by removal of UDP and epimerization at C2 by a hydrolase/epimerase enzyme (PtmD). This *manno* intermediate would then be condensed with P-enolpyruvate by a synthase (PtmC) to produce legionaminic acid and activated with a CMP-synthetase (PtmB) to generate CMP-Leg5Ac7Ac (Fig. 7 and Table 2). In a similar fashion to the *pse* pathway, further modification to CMP-Leg5Am7Ac would occur by the action of PtmG, a homolog of Cj1316 (*pseA*). CMP-Leg5AmNMe7Ac would then be produced through the action of an *N*-methyltransferase, PtmH. Interestingly, preliminary metabolomic analysis of a *pglE* mutant, which is the enzyme responsible for the second step in the biosynthesis of UDP-QuiNAc4NAc from UDP-GlcNAc, revealed no effect on the production of CMP-Leg nucleotides. Although this finding suggests that PglE may not be involved in making Leg sugars, the precise enzymatic steps of the legionaminic acid pathway in *Campylobacter* are currently under investigation.

DISCUSSION

The current study elucidates the structure of a second major flagellar glycan modification found on *Campylobacter* flagellins following purification of the nucleotide-activated metabolites

by HILIC-MS and NMR structural analysis. In addition to biosynthesis of CMP-Pse5Ac7Ac, *C. coli* VC167 also produces CMP-Leg5Am7Ac and CMP-Leg5AmNMe7Ac. These Leg monosaccharides had only been identified hitherto as components of lipopolysaccharide from a number of bacterial pathogens, including *L. pneumophila* and *Pseudomonas aeruginosa* (for review see Ref. 30). The structural differences between these sugars are shown in Fig. 8. In *L. pneumophila* subtle changes in structure of lipopolysaccharide sugars can dictate a highly specific immune response (43). It remains to be established if the considerable glycan biosynthetic potential among *Campylobacter* isolates reflects a structural variability that leads to a unique antigenic response during infection. Additionally, the surface localization of these glycan moieties on the assembled flagellar filament may also contribute to unique properties in terms of overall surface charge and be of significant relevance in host-pathogen interactions (15).

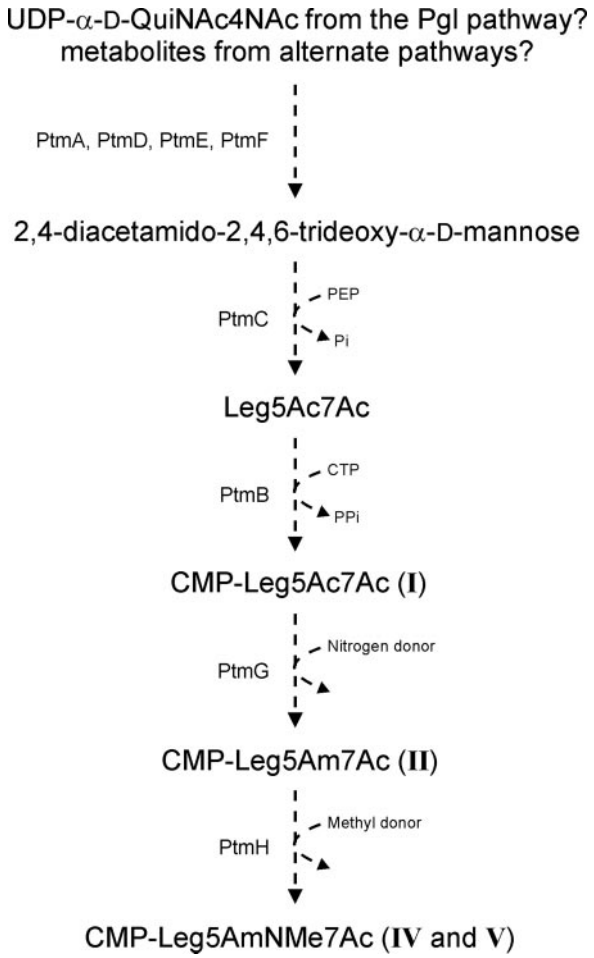

FIGURE 7. **Putative biosynthetic pathway for CMP-Leg derivatives found on *C. coli* VC167 flagellin.** Presumed activities are based on homology to pseudaminic acid biosynthetic enzymes PseI (PtmC), PseF (PtmB), and PseA (PtmG) as described in a previous study (20). Chemical structures of products **II**, **IV**, and **V** are shown in Figs. 4 and 5. The *broken arrows* indicate that the precise functional characterization of each biosynthetic step has yet to be determined.

The current study has also identified two previously unrecognized genes in the Leg pathway, *ptmG* and *ptmH*. Through targeted metabolomics analyses we have also been able to characterize further eight genes involved in the Leg biosynthetic pathway and provide tentative functional assignments for a number of gene products (Fig. 7). As was the case with pseudaminic acid until very recently, the biosynthetic pathway of legionaminic acid is currently unknown. A 30-kb lipopolysaccharide locus has been identified in *L. pneumophila* that appears to contain the legionaminic acid biosynthetic genes, although the functional analysis of respective gene products has yet to be determined (44). Putative functions were assigned for a number of these genes based on homology of the encoded proteins to the sialic acid biosynthetic enzymes NeuB (PtmC), NeuA (PtmB), and NeuC (PtmD). One of the more noteworthy revelations from the genomic sequence of *C. jejuni* NCTC11168 was the presence of three sets of genes that encoded proteins with similarity to sialic acid biosynthetic enzymes. It has been clearly established that one set of these genes is responsible for the biosynthesis of sialic acid, which is incorporated into the lipooligosaccharide (45), whereas the second set (*pseI* and *-F*) was recently shown to be responsible for key steps in the biosynthesis of pseudaminic acid (20), the novel flagellar glycan modification. We now show

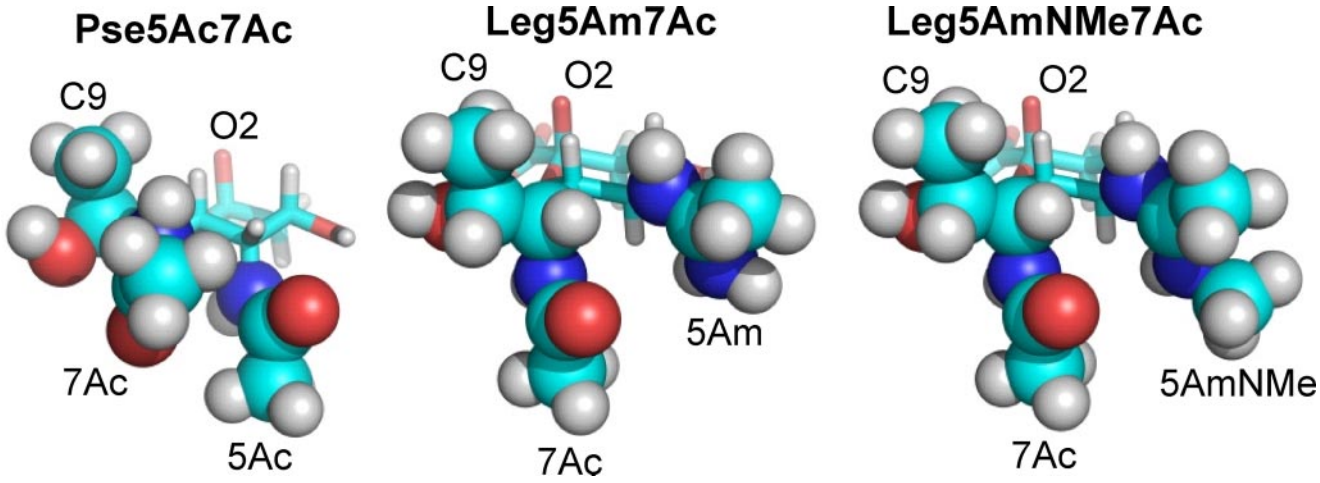

FIGURE 8. **Molecular models of Pse5Ac7Ac, Leg5Am7Ac, and Leg5AmNMe7Ac.** The exocyclic chain (C7–C9) and pendant groups at C5 and C7 are displayed as reduced van der Walls spheres to show the extent of their structural diversities. PDB files are available at ibs-isb.nrc-cnrc.gc.ca/facilities/NMR/molecularmodelling-e.html.

conclusively that the third set of these sialic acid biosynthetic enzyme homologs is part of a pathway responsible for the biosynthesis of the structurally related nonulosonate, legionaminic acid.

In the *L. pneumophila* work it was speculated that a derivative of *N*-acetylmannosamine would be the biosynthetic precursor for legionaminic acid in an analogous fashion to sialic acid (44). However, because not all *Campylobacter* strains have the genetic potential to synthesize Neu5Ac (17, 46)[4] but do have the ability to synthesize UDP-QuiNAc4NAc as a component of the *pgl N*-linked glycan, we felt that this intermediate was a logical choice as initial starting material for legionaminic acid biosynthesis. However, our current analysis indicates that this may not be the case. Inactivation of the *pgl* biosynthetic pathway enzyme PglE, which would prevent the production of all but the first biosynthetic precursors of this pathway, including UDP-QuiNAc4NAc, appeared to have no effect on the ability of *C. coli* VC167 to produce CMP-Leg nucleotides. It therefore appears that a distinct pathway producing novel intermediates is required for the biosynthesis of legionaminic acids in *Campylobacter*.

Based on significant homology to the *pse* pathway enzymes, we can now tentatively assign the specific functions of *ptmC*, *ptmB*, *ptmG*, and *ptmH* as encoding a Leg5Ac7Ac synthase, a CMP-Leg5Ac7Ac synthetase, a CMP-Leg5Am7Ac acetamidino synthase, and an acetamidino *N*-methyltransferase, respectively (Fig. 7). The precise biosynthetic route taken to produce the final two novel glycan modifications and the roles of *ptmA*, *-D*, *-E*, and *-F* in the Leg biosynthetic pathway remain to be established. Also, there may be additional genes in the Leg pathway that have yet to be identified.

In contrast to studies in *C. jejuni* 81–176 where we were able to demonstrate the accumulation of novel UDP-linked sugar nucleotides in the cell metabolome when the *pse* pathway was interrupted, no such accumulation was evident in mutants of legionaminic acid biosynthetic genes in *C. coli* VC167. Although accumulation of intermediates was shown to be due to cross-talk between the *pse* and *pgl* pathways in *C. jejuni* 81–176 (19), we did not observe cross-talk for the Leg pathway. This may be a reflection of the rapid utilization of intermediates of the Leg pathway in alternate metabolic pathways of the cell.

This study reveals for the first time that most *Campylobacter* isolates have the genetic potential to synthesize legionaminic acids in addition to pseudaminic acids. We have shown in this study and a previous one (11), that these can also be incorporated into the flagellar filament glycoprotein, flagellin. The potential to exploit the enzymes from these pathways in glycoengineering applications is significant, especially because legionaminic acid is an analog of sialic acid (Neu5Ac) with the same D-*glycero*-D-*galacto* absolute configuration. The targeted metabolomics approach used in this study has revealed a number of genetic targets encoding biosynthetic enzymes whose precise function can now be verified through recombinant expression and *in vitro* functional assays.

[4] F. Poly, T. Read, and P. Guerry, unpublished observation.

Acknowledgments—We thank Michael Schirm for the top down analysis of the flagellin protein and Ian Schoenhofen for discussion and input.

REFERENCES

1. Blaser, M. J. (1997) *J. Infect. Dis.* **176,** Suppl. 2, S103–S105
2. Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M., and Tauxe, R. V. (1999) *Emerg. Infect. Dis.* **5,** 607–625
3. Konkel, M. E., Klena, J. D., Rivera-Amill, V., Monteville, M. R., Biswas, D., Raphael, B., and Mickelson, J. (2004) *J. Bacteriol.* **186,** 3296–3303
4. Song, Y. C., Jin, S., Louie, H., Ng, D., Lau, R., Zhang, Y., Weerasekera, R., Al Rashid, S., Ward, L. A., Der, S. D., and Chan, V. L. (2004) *Mol. Microbiol.* **53,** 541–553
5. Black, R. E., Levine, M. M., Clements, M. L., Hughes, T. P., and Blaser, M. J. (1988) *J. Infect. Dis.* **157,** 472–479
6. Blaser, M. J., and Duncan, D. J. (1984) *Infect. Immun.* **44,** 292–298
7. Martin, P. M., Mathiot, J., Ipero, J., Kirimat, M., Georges, A. J., and Georges-Courbot, M. C. (1989) *Infect. Immun.* **57,** 2542–2546
8. Nachamkin, I., and Hart, A. M. (1985) *J. Clin. Microbiol.* **21,** 33–38
9. Guerry, P., Alm, R. A., Power, M. E., Logan, S. M., and Trust, T. J. (1991) *J. Bacteriol.* **173,** 4757–4764
10. Thibault, P., Logan, S. M., Kelly, J. F., Brisson, J. R., Ewing, C. P., Trust, T. J., and Guerry, P. (2001) *J. Biol. Chem.* **276,** 34862–34870
11. Logan, S. M., Kelly, J. F., Thibault, P., Ewing, C. P., and Guerry, P. (2002) *Mol. Microbiol.* **46,** 587–597
12. Doig, P., Kinsella, N., Guerry, P., and Trust, T. J. (1996) *Mol. Microbiol.* **19,** 379–387
13. Power, M. E., Guerry, P., McCubbin, W. D., Kay, C. M., and Trust, T. J. (1994) *J. Bacteriol.* **176,** 3303–3313
14. Alm, R. A., Guerry, P., Power, M. E., and Trust, T. J. (1992) *J. Bacteriol.* **174,** 4230–4238
15. Guerry, P., Ewing, C. P., Schirm, M., Lorenzo, M., Kelly, J., Pattarini, D., Majam, G., Thibault, P., and Logan, S. M. (2006) *Mol. Microbiol.* **60,** 299–311
16. 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., Rajandream, M. A., Rutherford, K. M., van Vliet, A. H., Whitehead, S., and Barrell, B. G. (2000) *Nature* **403,** 665–668
17. Fouts, D. E., Mongodin, E. F., Mandrell, R. E., Miller, W. G., Rasko, D. A., Ravel, J., Brinkac, L. M., Deboy, R. T., Parker, C. T., Daugherty, S. C., Dodson, R. J., Durkin, A. S., Madupu, R., Sullivan, S. A., Shetty, J. U., Ayodeji, M. A., Shvartsbeyn, A., Schatz, M. C., Badger, J. H., Fraser, C. M., and Nelson, K. E. (2005) *PLoS. Biol.* **3,** e15
18. McNally, D. J., Hui, J. P., Aubry, A. J., Mui, K. K., Guerry, P., Brisson, J. R., Logan, S. M., and Soo, E. C. (2006) *J. Biol. Chem.* **281,** 18489–18498
19. Schoenhofen, I. C., McNally, D. J., Vinogradov, E., Whitfield, D., Young, M., Dick, S., Wakarchuk, W. W., Brisson, J.-R., and Logan, S. M. (2006) *J. Biol. Chem.* **281,** 723–732
20. Schoenhofen, I. C., McNally, D. J., Brisson, J. R., and Logan, S. M. (2006) *Glycobiology* **16,** 8C–14C
21. Soo, E. C., Aubry, A. J., Logan, S. M., Guerry, P., Kelly, J. F., Young, N. M., and Thibault, P. (2004) *Anal. Chem.* **76,** 619–626
22. Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003) *Mol. Microbiol.* **48,** 1579–1592
23. Harris, L. A., Logan, S. M., Guerry, P., and Trust, T. J. (1987) *J. Bacteriol.* **169,** 5066–5071
24. Labigne-Roussel, A., Courcoux, P., and Tompkins, L. (1988) *J. Bacteriol.* **170,** 1704–1708
25. Brisson, J. R., Sue, S. C., Wu, W. G., McManus, G., Nghia, P. T., and Uhrin, D. (2002) in *NMR Spectroscopy of Glycoconjugates* (Jimenez-Barbero, J., and Peters, T., eds) pp. 59–93, Wiley-VCH, Weinhem, Germany
26. Brisson, J. R., Crawford, E., Khieu, N. H., Perry, M. B., and Richards, J. C. (2002) *Can. J. Chem.* **80,** 949–963
27. Uhrin, D., and Brisson, J.-R. (2000) in *NMR in Microbiology: Theory and Applications* (Barotin, J. N., and Portais, J. C., eds) pp. 165–210, Horizon

Scientific Press, Wymonden, UK
28. Schirm, M., Schoenhofen, I. C., Logan, S. M., Waldron, K. C., and Thibault, P. (2005) *Anal. Chem.* **77,** 7774–7782
29. Knirel, Y. A., Kocharova, N. A., Shashkov, A. S., Dmitriev, B. A., Kochetkov, N. K., Stanislavsky, E. S., and Mashilova, G. M. (1987) *Eur. J. Biochem.* **163,** 639–652
30. Knirel, Y. A., Shashkov, A. S., Tsvetkov, Y. E., Jansson, P. E., and Zahringer, U. (2003) *Adv. Carbohydr. Chem. Biochem.* **58,** 371–417
31. Tsvetkov, Y. E., Shashkov, A. S., Knirel, Y. A., and Zahringer, U. (2001) *Carbohydr. Res.* **331,** 233–237
32. Vliegenthart, A. (1982) in *Cell Biology Monographs* (Schauer, R., ed) pp. 127–172, Springer-Verlag, NewYork
33. Tsvetkov, Y. E., Shashkov, A. S., Knirel, Y. A., and Zahringer, U. (2001) *Carbohydr. Res.* **335,** 221–243
34. Hermansson, K., Perry, M. B., Altman, E., Brisson, J. R., and Garcia, M. M. (1993) *Eur. J. Biochem.* **212,** 801–809
35. Kenne, L., Lindberg, B., Schweda, E., Gustafsson, B., and Holme, T. (1998) *Carbohydr. Res.* **180,** 285–294
36. Kooistra, O., Luneberg, E., Knirel, Y. A., Frosch, M., and Zahringer, U. (2002) *Eur. J. Biochem.* **269,** 560–572
37. Kooistra, O., Herfurth, L., Luneberg, E., Frosch, M., Peters, T., and Zahringer, U. (2002) *Eur. J. Biochem.* **269,** 573–582
38. Lindberg, B. (1990) *Adv. Carbohydr. Chem. Biochem.* **48,** 279–318
39. Yao, R., Alm, R. A., Trust, T. J., and Guerry, P. (1993) *Gene* (*Amst.*) **130,** 127–130
40. Goon, S., Kelly, J. F., Logan, S. M., Ewing, C. P., and Guerry, P. (2003) *Mol. Microbiol.* **50,** 659–671
41. Young, N. M., Brisson, J. R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C. M. (2002) *J. Biol. Chem.* **277,** 42530–42539
42. Oliver, N. B., Chen, M. M., Behr, J. R., and Imperiali, B. (2006) *Biochemistry* **45,** 13659–13669
43. Luneberg, E., Zahringer, U., Knirel, Y. A., Steinmann, D., Hartmann, M., Steinmetz, I., Rohde, M., Kohl, J., and Frosch, M. (1998) *J. Exp. Med.* **188,** 49–60
44. Luneberg, E., Zetzmann, N., Alber, D., Knirel, Y. A., Kooistra, O., Zahringer, U., and Frosch, M. (2000) *Int. J. Med. Microbiol.* **290,** 37–49
45. 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
46. Aspinall, G. O., Lynch, C. M., Pang, H., Shaver, R. T., and Moran, A. P. (1995) *Eur. J. Biochem.* **231,** 570–578
47. Guerry, P., Doig, P., Alm, R. A., Burr, D., Kinsella, N., and Trust, T. J. (1996) *Mol. Microbiol.* **19,** 369–378

Targeted Metabolomics Analysis of *Campylobacter coli* VC167 Reveals Legionaminic Acid Derivatives as Novel Flagellar Glycans

David J. McNally, Annie J. Aubry, Joseph P. M. Hui, Nam H. Khieu, Dennis Whitfield, Cheryl P. Ewing, Patricia Guerry, Jean-Robert Brisson, Susan M. Logan and Evelyn C. Soo

J. Biol. Chem. 2007, 282:14463-14475.
doi: 10.1074/jbc.M611027200 originally published online March 19, 2007
