Anal. Chem. **2005,** *77,* 4897−4907

Genomic and Proteomic Identification of a DNA-Binding Protein Used in the "Fingerprinting" of *Campylobacter* Species and Strains by MALDI-TOF-MS Protein Biomarker Analysis

Clifton K. Fagerquist,* William G. Miller, Leslie A. Harden, Anna H. Bates, William H. Vensel, Guilin Wang, and Robert E. Mandrell

Western Regional Research Service, Agricultural Research Service, U.S. Department of Agriculture, 800 Buchanan Street, Albany, California 94710

We have identified a prominent ~10-kDa protein biomarker observed in the matrix-assisted laser desorption/ionization time-of-flight mass spectra (MALDI-TOF-MS) of cell lysates of five thermophilic species of *Campylobacter*: *jejuni, coli, lari, upsaliensis*, and *helveticus*. The biomarker was unambiguously identified by genomic and proteomic sequencing as a DNA-binding protein HU. We report the amino acid sequence of HU as determined by sequencing the *hup* gene of four species (12 strains): *C. jejuni* (2), *C. coli* (4), *C. upsaliensis* (4) and *C. lari* (2). Confirmation of the amino acid sequence was obtained by nanoflow high-performance liquid chromatography-tandem mass spectrometry of the tryptic peptides of the extracted/digested HU protein. Protein identification was also confirmed by comparison of the molecular weight (MW) predicted from the *hup* gene and the MW of HU as measured by high-resolution mass spectrometry. We found the HU protein to be particularly useful as a biomarker in that it strongly ionizes by MALDI and its MW varies *between* species and among strains *within* a species. Intra- and interspecies variation of the HU MW is due to changes in the amino acid sequence of the HU protein and not due to co- or posttranslational modifications. The strong ionization efficiency of HU by MALDI is likely due, in part, to four lysine residues clustered at the carboxyl end of the protein. We also report identification of the HU protein biomarker for a *C. helveticus* strain, whose *hup* gene was not sequenced, but whose HU amino acid sequence was partially conserved in *C. upsaliensis* strains. We have also tentatively assigned a ~10.5-kDa protein biomarker of a *C. concisus* strain as an HU protein.

The rapid detection and analysis of pathogens is an increasingly important area of analytical science. Research in this area has been spurred by recent concerns about biosecurity and the possible use of pathogens as threat agents in a terrorist attack. However, there are a number of areas where pathogen detection and characterization has always been important, such as food safety, public health, and clinical medicine. Sensitivity, specificity, speed, simplicity, and ruggedness are all critical objectives in the improvement of existing methods for detection as well as the development of new techniques. Mass spectrometry has emerged as a powerful tool for the detection and analysis of pathogens and a relatively simple technique for the characterization of microorganisms.[1] Early studies utilizing mass spectrometry for bacterial identification analyzed small-molecule biomarkers, e.g., lipids, phospholipids, and carbohydrates.[2,3] As new techniques for ionization and mass analysis developed in the 1980s and 1990s, the ability to detect and characterize pathogens on the basis of larger biomolecules expanded dramatically. The most significant of these techniques was the coupling of matrix-assisted laser desorption/ionization (MALDI) with time-of-flight mass spectrometry (TOF-MS).[4,5] A number of early studies proved the feasibility of bacterial identification using MALDI-TOF-MS analysis of protein biomarkers from protein extracts of cell lysates or from whole cells.[6−14]

* To whom correspondence should be addressed. E-mail: cfagerquist@pw.usda.gov.

(1) Anhalt, J.; Fenselau, C. *Anal. Chem.* **1975**, *47*, 219−225.
(2) Heller, D.; Fenselau, C.; Cotter, R.; Demirev, P.; Olthoff, J.; Honovich, J.; Uy, M.; Tanaka, T.; Kishimoto, Y. *Biochem. Biophys. Res. Commun*. **1987**, *142*, 194−199.
(3) *Mass Spectrometry for the Characterization of Microorganisms*; Fenselau, C., Ed.; ACS Symposium Series 541; American Chemical Society: Washington, DC, 1994.
(4) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom*. **1988**, *2*, 151−153.
(5) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem*. **1991**, *63*, 1193A−1202A.
(6) Cain, T. C.; Lubman, D. M.; Weber, W. J., Jr. *Rapid Commun. Mass Spectrom*. **1994**, *8*, 1026−1030.
(7) Krishnamurthy, T.; Ross, P. L.; Rajamani, U. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 883−888.
(8) Krishnamurthy, T.; Ross, P. L. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1992−1996.
(9) Holland, R. D.; Wilkes, J. G.; Rafii, F.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O., Jr. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1227−1232.
(10) Arnold, R.; Reilly, J. *Rapid Commun. Mass Spectrom*. **1998**, *12*, 630−636.
(11) Welham, K.; Domin, M.; Scannell, D.; Cohen, E.; Ashton, D. *Rapid Commun. Mass Spectrom*. **1998**, *12*, 176−180.
(12) Haag, A.; Taylor, S.; Johnston, K.; Cole, R. *J. Mass Spectrom*. **1998**, *33*, 750−756.
(13) Wang, Z.; Russon, L.; Li, L.; Roser, D.; Long, S. R. *Rapid Commun. Mass Spectrom*. **1998**, *12*, 456−464.
(14) Dai, Y.; Li, L.; Roser, D.; Long, S. R. *Rapid Commun. Mass Spectrom*. **1999**, *13*, 73−78.

10.1021/ac040193z Not subject to U.S. Copyright. Publ. 2005 Am. Chem. Soc.
Published on Web 06/24/2005

Characterization of pathogens on the basis of their protein biomarkers had the advantage that the composition and sequence of the biomarker was encoded in the organism's genome.[15,16]

Two basic approaches have emerged for the identification of bacteria (and other microorganisms) by MALDI-TOF-MS analysis. The first involves comparison of the mass spectrum of a bacterial sample of unknown genus/species/strain to that of a bacterial sample of known genus/species/strain. By this approach, a comparison is made of the mass-to-charge (m/z) of prominent protein biomarkers (usually 10−30) in the known and unknown sample without concern for the actual identity of the biomarkers themselves. Due to variations in biomarker ion signal intensity from sample to sample, the relative intensity of the biomarker ions has not been found to be a reliable criterion for bacterial identification. Various algorithms have been developed by a number of research groups (including ours) to compare mass spectral profiles of known and unknown samples.[17,18] This mass spectral "fingerprint" approach involves the use of statistical methods, pattern recognition algorithms, etc., to identify similarities between the m/z of biomarker ions of an unknown sample and biomarker ions from a library of known reference samples.

The second approach may be described as a bioinformatics approach that exploits genomic databases.[19−23] Specifically, the predicted molecular weights (MWs) of proteins encoded by microbial genomes are compared to the m/z of the biomarker ions in the MALDI-TOF-MS spectra of an unknown pathogen. The pathogen whose genome has the greatest number of protein MWs that correspond to observed biomarkers provides identification of the pathogen and concomitant assignment of the protein biomarkers. Demirev, Fenselau, and co-workers have pioneered this approach and developed algorithms that also incorporate potential posttranslational modifications of bacterial proteins, e.g., *N*-terminal methionine cleavage.[21] One advantage of this identification strategy is that as more microbial genomes become available, the strength of this approach becomes more powerful.

Among human microbial pathogens, of particular importance is *Campylobacter*, which has been estimated to cause ~2.4 million incidents of food-borne illness every year in the United States alone, in addition to being a significant health problem worldwide.[24] Of the 16 species of *Campylobacter*, 11 are potential human pathogens: *C. jejuni, C. coli, C. lari, C. upsaliensis, C. sputorum*, *C. concisus*, *C. fetus, C. gracilis*, *C. showae*, *C. rectus*, and *C. hyointestinalis*. *C. jejuni* is responsible for 99% of the illnesses;[24] however, the remaining pathogenic species have been shown increasingly to be responsible for human illness, especially in the developing world.[25] MALDI-TOF-MS analysis of *Campylobacter* has been limited. Winkler et al. reported identification of discriminatory protein biomarker ions of *C. jejuni*, *C. fetus*, and *C. coli* as well as *Helicobacter pylori* and *Helicobacter mustelae* using MALDI-TOF-MS.[26] Very recently, Mandrell, Harden, and co-workers reported using MALDI-TOF-MS to identify a number of species and strains of *Campylobacter* from their protein biomarkers. Tentative assignment of several biomarkers was accomplished by comparing biomarker ion m/z to predicted MWs of proteins from an in-house database of *Campylobacter* proteins generated from genomic sequencing. Preliminary peptide mass mapping results also indicated, in a few strains, the identity of a prominent biomarker as that of a DNA-binding protein HU.[27]

In the current study, using more extensive genomic and proteomic techniques, we confirm and extend identification of the DNA-binding protein HU as a biomarker that is consistently prominent in the MALDI-TOF-MS spectra of cell lysates of several species and strains of *Campylobacter*. In contrast, with other protein biomarkers used for microbial identification, the HU protein does not appear to undergo posttranslational modifications (PTMs) common to bacterial proteins, e.g., N-terminal methionine cleavage. In consequence, there is a direct correspondence between the genomically determined amino acid sequence and the sequence of the observed biomarker. Although labor intensive, definitive identification of protein biomarkers by genomic/proteomic techniques provides increased information about species and strain differences within a genus. In addition, such work provides further validation and support for the use of MALDI-TOF protein biomarker analysis for the rapid identification of pathogens. Parts of this work were previously presented.[28]

MATERIALS AND METHODS

Materials. All chemicals and solvents were purchased from either Sigma-Aldrich (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA, and Fair Lawn, NJ) unless otherwise noted. Gels, buffers, standards and electrophoresis apparatus for protein separation were purchased from Invitrogen (Carlsbad, CA,). Gases were obtained from Praxair (Oakland, CA). PCR enzymes and reagents were purchased from New England Biolabs (Beverly, MA) or Epicenter (Madison, WI). DNA sequencing chemicals and capillaries were purchased from Applied Biosystems (Foster City, CA).

Bacterial Culture. Growing *Campylobacte*r was described previously.[27] *Warning*: *Campylobacter* is a Biosafety Level 2 human pathogen. All appropriate precautions were taken when handling this pathogen. Briefly, all *Campylobacter* strains were kept frozen at −80 °C in BioBank beads (Microbank, Ontario, Canada) until inoculation on nutrient agar. *Campylobacter ssp. jejuni* and *coli* were grown at 37−42 °C on Brucella Broth Agar (BBA, Brucella Broth BBL, Becton Dickinson, Sparks, MD) containing 1.5% Bacto Agar (Becton Dickinson) and supplemented with 0.25 g/L sodium metabisulfite, 0.25 g/L sodium pyruvate, and 0.25 g/L ferrous

(15) Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001**, *20*, 157−171.
(16) Lay, J. O., Jr. *Mass Spectrom. Rev.* **2001**, *20*, 172−194.
(17) Jarmon, K. H.; Cebula, S. T.; Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Kingsley, M. T.; Wahl, K. L. *Anal. Chem.* **2000**, *72*, 1217−1223.
(18) Wahl, K. L.; Wunschel, S. C.; Jarman, K. H.; Valentine, N. B.; Petersen, C. E.; Kingsley, M. T.; Zartolas, K. A.; Saenz, A. J. *Anal. Chem.* **2002**, *74*, 6191−6199.
(19) Demirev, P. A.; Ho, Y.-P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 2732−2738.
(20) Peneda, F. J.; Lin, J. S.; Fenselau, C.; Demirev, P. A. *Anal. Chem.* **2000**, *72*, 3739−3744.
(21) Demirev, P. A.; Lin, J. S.; Peneda, F. J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 4566−4573.
(22) Yao, Z.-P.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2002**, *74*, 2529−2534.
(23) Peneda, F. J.; Antoine, M. D.; Demirev, P. A.; Feldman, A. B.; Jackman, J.; Longenecker, M.; Lin, J. S. *Anal. Chem.* **2003**, *75*, 3817−3822.
(24) http://www.cdc.gov/ncidod/dbmd/diseaseinfo/campylobacter_t.htm.
(25) Lastovica, A. J.; le Roux, E. *J. Clin. Microbiol.* **2000**, *38*, 2798−2799.
(26) Winkler M. A.; Uher, J.; Cepa, S. *Anal. Chem.* **1999**, *71*, 3416−3419.
(27) Mandrell, R. E.; Harden, L. A.; Bates, A. H.; Miller, W. G.; Haddon, W. F.; Fagerquist, C. K. Submitted.
(28) Fagerquist, C. K.; Harden, L. A.; Miller, W. G.; Fouts, D. E.; Nelson, K. E.; Mandrell, R. E. *52nd ASMS Conf. Mass Spectrom. Allied Top.*, Nashville, TN, May 23−27, 2004.

sulfate. Brain heart infusion medium was used instead of BBA when the objective was DNA extraction. *Campylobacter ssp. upsaliensis, lari*, and *helveticus* were grown at 37 °C on BBA supplemented with 50−100 mL of laked horse blood (BAB, Hema Resources, Aurora, OR) per liter of agar. *Campylobacter ssp. concisus* was grown at 37 °C on BBA supplemented with laked horse blood, and 0.20 g/L L-cysteine, 0.60 g/L sodium formate, and 0.60 g/L sodium fumarate. The inoculated agar plates were placed in plastic Zip-Lock freezer bags. The bags were filled/purged three times with a compressed certified gas mixture of 5−10% hydrogen, 10% carbon dioxide, and 80−85% nitrogen and then given a final fill. Additional hydrogen gas and reduced oxygen was provided for the more fastidious *C. concisus* strains and certain *C. upsaliensis* strains by addition of a CampyPak sachet (BBL) to the Zip-Lock bag prior to infusing the bag with the gas mixture. After 48 h of growth (72 h for *C. concisus*), the bacteria were harvested for MALDI-TOF-MS biomarker analysis, protein extraction, or DNA extraction.

DNA Preparation. Extraction of *Campylobacter* DNA was described previously.[29] Briefly, cells were harvested with a 10-μL inoculating loop from agar plates and resuspended in 1.5 mL of 10% (w/v) sucrose, 50 mM Tris (pH 8.0). A 250-μL aliquot of a 10 mg/mL lysozyme solution (in 250 mM Tris, pH 8.0) and 600 μL of 0.1 M EDTA were then added to the suspension. The suspension was incubated for 10 min on ice, then 300 μL of a 5% (w/v) SDS solution was added, and the mixture was vortexed briefly to clarify the solution. The lysates were incubated sequentially with RNAase (1 mg/mL) and proteinase K (10 mg/mL), and the DNA was spooled onto a hood-ended glass Pasteur pipet following addition of sodium acetate (1/10 volume) and ethanol (2 volumes). DNA was resuspended in Tris−EDTA (pH 8.0), extracted twice with phenol/chloroform (1:1 v/v), once with chloroform, and concentrated by ethanol precipitation.

Amplification and Sequencing of the *hup* Gene. The *C. jejuni* and *C. coli hup* genes were amplified with the cjchupF (5′ AAT GAA ATT GAY AAT AGT GGR GAA T 3′) and cjchupR (5′ TTT CCT ATA AGY TCA YTT ACT TTT T 3′) primer sets. The *C. lari hup* genes were amplified with the larihupfullF (5′ TAA AGC CTG AAG TGG TAA TAA ATC TAG 3′) and larihupfullR (5′ ATA ATC GGG GTA TTA CCT ATA CAA TCT 3′) primer sets. The *C. upsaliensis hup* genes were amplified with the cuhupF (5′ GTG GAG AAT TTG AAC GCT ACA CT 3′) and cuhupR (5′ TAA TGA GCT TTA CCT GCT TGA TG 3′) primer sets. PCR reactions were performed on an MJ Research (San Francisco, CA) Tetrad thermocycler with the following settings: 30 s at 94 °C; 30 s at 53 °C; 2 min at 72 °C (30 cycles). Each amplification mixture contained 50 ng of genomic DNA, 1 × PCR buffer (Epicenter), 1 × PCR enhancer (Epicenter), 250 μM $MgCl_2$, 250 μM each dNTP, 50 pmol of each primer, and 0.2 unit of polymerase (New England Biolabs). Amplicons were purified on a Qiagen BioRobot 8000 workstation (Qiagen, Santa Clarita, CA).

Cycle sequencing reactions were performed on an MJ Research Tetrad thermocycler using the ABI PRISM BigDye terminator cycle sequencing kit (Version 3.0, Applied Biosystems, Foster City, CA) using standard protocols. All extension products were purified on DyeEx 96 well plates (Qiagen). DNA sequencing was performed on an ABI PRISM 3100 genetic analyzer using the POP-6 polymer and ABI PRISM Genetic Analyzer Data Collection and ABI PRISM Genetic Analyzer Sequencing Analysis software. PCR/sequencing oligonucleotides were purchased from Qiagen.

Protein Extraction. For extraction of the protein biomarker from cell lysates, a 10-μL disposable inoculating loop (NUNC Brand Products) was used to transfer cells from the culture plate to four extraction tubes each containing 0.5 mL of the same solvent mixture and beads used for MALDI-TOF analysis of cell lysates (see Mass Spectrometry: MALDI-TOF-MS section). After bead-beating and centrifugation, the supernatant of the tubes was transferred to a 15-mL polypropylene tube. The combined supernatant was ∼2 mL to which was added 3.0 mL of 0.1% trifluoroacetic acid (TFA). The tube was capped, briefly vortexed, and partitioned into two 2.0-mL Ultrafree-CL centrifugal filter devices (Durapore, PVDF membrane, 0.45 μm, Micron Bioseparations, Millipore, Bedford, MA). The filter devices were then centrifuged for 1 h at 4000 rpm at 4 °C using an automatic refrigerated centrifuge. After centrifugation, the filtrate was stored at −80 °C for later analysis by HPLC.

HPLC Separation. The cell lysate supernatant solution was chromatographically separated using a Hewlett-Packard (Palo Alto, CA) Series II 1090 liquid chromatograph. The solution was injected onto a 5.0-mL loop using a gastight 5.0 mL Hamilton (Reno, NV) glass syringe fitted with a blunt-tipped needle. The column used for chromatographic separation was a large-capacity protein and peptide C18 Vydac column. HP Chemstation software was used to operate the LC and acquire data. Proteins, peptides, and other cell lysate analytes were detected by UV absorbance (λ = 210 nm). The column oven was heated to 50 °C. Mobile phase A was 0.1% TFA in HPLC grade water. Mobile phase B was 1:1 acetonitrile/2-propanol with 0.1% TFA. The flow rate was 1.5 mL/min. The HPLC gradient was as follows: 0−5 (85% A), 5−65 (85−20% A), 65−80 (20% A), and 80−90 min (20−90% A). Eluent was collected manually in 2-mL snap-cap Eppendorfs. About ∼90 HPLC fractions were collected over a period of ∼1 h. The protein biomarker, subsequently identified as HU, consistently showed a prominent peak in the HPLC chromatogram (RT ∼ 40−50 min) suggesting a highly expressed protein. HPLC fractions were analyzed by MALDI-TOF analysis (or stored at −80 °C for later analysis).

1-D SDS−PAGE of HPLC Fractions. Using a centrifugal evaporator, 0.5−1.0 mL of HPLC eluent was evaporated from the fraction containing the protein biomarker (as identified by MALDI-TOF-MS) to give a final volume of 20−30 μL. To this volume was added 10 μL of Nupage LDS sample buffer (Invitrogen). The sample was allowed to react for 1.5 h at room temperature followed by 10 min at 65 °C. The 30−35-μL sample was loaded onto a 4−12% Bis-Tris gel, 1.5 mm × 10 well. Prior to sample loading, the gel wells were prepared according to manufacturer's instructions using a 20-fold dilution of Nupage MES SDA running buffer. The gel cassette was run in a Novex minicell (Xcell SureLock). Protein standards (Mark12) occupied lanes 1 and 10. Electrophoresis was performed using a constant-voltage (200 V) program for a duration of 35−40 min. After electrophoresis, the gel was stained for 48 h with SYPRO Ruby protein gel stain solution (Molecular Probes, Eugene, OR). The gels were destained with a solution of 10% HPLC grade methanol, 7% reagent grade glacial

(29) Miller, W. G.; On, S. L. W.; Wang, G.; Fontanoz, S.; Lastovica, A. J.; Mandrell, R. E. *J. Clin. Microbiol.* In press.

acetic acid, and 83% HPLC grade water by volume. Gels were photographed with a gel/blot imager (Alpha Innotech, San Leandro, CA) using trans-illumination at a wavelength of 302 nm.

In-Gel Digestion of Protein Biomarker. Gel bands were excised while illuminated on a UV light box (Fotodyne Inc., Hartland, WI). *Cautionary note*: UV radiation is harmful to eyes and skin. Precautions were taken to cover all exposed skin. In addition, both UV-resistant glasses and a UV-resistant shield were worn for eye protection. The excised gel bands were further divided to produce ~1-mm^3 gel cubes. Two (or three) gel cubes of each excised band were deposited into a well of the 96-well ProGest processing tray (Genomic Solutions, Ann Arbor, MI). Two holes in the base of each well of the processing tray allow for transfer of reagents to waste or digestion product to the receiving tray under a pressure of 7.0 psi nitrogen (99.999% purity). In-gel digestion was performed robotically on a DigestPro automated protein digester (Intavis, Langenfeld, Germany) according to manufacturer's instructions. The five (or six) of gel cubes obtained from a single band allowed for multiple digestion experiments (if necessary). Porcine trypsin (Promega, Madison, WI) was used for enzymatic digestion. Tryptic peptides were eluted from the gel using 10% formic acid that was 0.1% in TFA and deposited into the 96-well receiving tray. The solution volume generated was typically 20−40 μL. Samples were then analyzed by nano-LC/MS/MS (or stored at −80 °C for later analysis).

Mass Spectrometry: MALDI-TOF-MS. (a) Cell Lysates. MALDI-TOF biomarker analysis of *Campylobacter* cell lysates has been described in detail elsewhere.[27] Briefly, *Campylobacter* cells were transferred using a 1-μL inoculating loop to a 2.0-mL microtube. The microtube contained 0.5 mL of water/acetonitrile (both HPLC grade) at a 2:1 ratio with 0.1% TFA (sequencing grade) and ~40 mg of 0.1 mm zirconia/silica beads (BioSpec Products Inc., Bartlesville, OK). The tube was then capped, placed into a 96-well aluminum holder, and agitated for 60 s on a beadbeater (Mini-beadbeater-96+, BioSpec Products, Bartlesville, OK) followed by centrifugation for 4−5 min at 10 000 rpm on a centrifuge (Eppendorf, model 4515C, Hamburg, Germany). A saturated solution of *trans*-4-hydroxy-3-methoxycinnamic acid (ferulic acid) was prepared in 2:1 water/acetonitrile with 0.1% TFA. A 100-μL aliquot of the saturated solution was diluted to 300 μL and then deposited in 0.5-μL aliquots onto a homemade stainless steel MALDI target plate with a 7 × 7 array spot locations. The matrix spots were allowed to dry at room temperature after which 0.5 μL of cell lysate supernatant was deposited onto the dried matrix spots. The spots were again allowed to dry at room temperature. The calibrants, myoglobin, angiotensin II, and adrenocorticotropic hormone human fragment 18−39 (ACTH Clip), were deposited at three corner spots of the target. The target plate was then introduced into the source of a Bruker Reflex II MALDI-TOF mass spectrometer (Billerica, MA). Laser desorption/ionization was achieved with a pulsed nitrogen laser (λ = 337 nm, 4-ns pulse width, 300 μJ/pulse, 2-Hz repetition rate). The instrument was externally calibrated in reflectron mode using delayed ion extraction at an ion acceleration voltage of 20 kV, which resulted in a mass accuracy of ±5−10 Da. Resolution was 800−1200 fwhm. Data were automatically or manually processed using Bruker XMASS software.

(b) HPLC Fractions. All ~90 HPLC fractions of each strain were analyzed by MALDI-TOF-MS in order to rapidly identify which fraction (or fractions) contained the putative HU protein biomarker. Analysis of the HPLC fractions by MALDI-TOF-MS was similar in procedure to analysis of cell lysates. A 0.5-μL aliquot of each HPLC fraction was spotted onto a dried spot of the ferulic acid matrix. For HPLC fractions of high organic content, a 0.5-μL aliquot of 0.2% TFA was spotted onto the target prior to deposition of the HPLC aliquot. Instrument operation and data acquisition were controlled by the Auto X tool of XTOF. We found that the MW of the HPLC protein biomarker and the MW of the cell lysate protein biomarker varied as much as ±5−10 Da in some samples due to differences in MALDI-TOF external calibration.

Mass Spectrometry: ESI-MS and MS/MS. (a) HPLC Fractions. Once identified by MALDI-TOF-MS, the tentative HPLC fraction containing HU was further analyzed by high-resolution electrospray ionization mass spectrometry (HR ESI-MS) using a quadrupole/time-of-flight mass spectrometer (Q-STAR Pulsar I, MDS Sciex/ABI, Toronto, Canada). Two fragment ions generated by the collision-activated dissociation of the +2 charge state of glufibrogen (m/z 785.8) were used to externally calibrate the reflectron time-of-flight mass analyzer of the Q-STAR. TOF resolution was typically ~8000−9000 fwhm. Approximately 8 μL of the HPLC fraction was loaded into an ESI spray capillary (NanoES, Proxeon Biosystems) using a 10-μL gel pipet tip. After sample loading and brief centrifugation, the spray capillary was mounted onto the source head and a backing pressure of 1−8 atm was applied from a 10-mL glass Luer-lock glass syringe. The spray voltage was 1800 V. The acquisition mass range was m/z 450−2000, and the accumulation time in the collision quadrupole was 1.0 s. The average MW of the protein biomarker was calculated by deconvolution of the charge-state envelope using the Bayesian protein reconstruct provided with the instrument software (Biotools, Analyst QS, Version 1.2, MDS Sciex/ABI). The average MW of the protein biomarker was measured three or four times with recalibration of TOF for each measurement. The predicted average MW of the HU biomarker was calculated from its genomic derived amino acid sequence using GPMAW software (Version 5.10, Lighthouse Data, Odense, Denmark).

(b) Nano-LC/MS/MS. Biomarker tryptic peptides were separated using a nanoflow high-performance liquid chromatograph system (LC Packings/Dionex, Sunnyvale, CA) interfaced to our quadrupole time-of-flight mass spectrometer. Samples were ionized by nanoelectrospray ionization (nano-ESI) using an 8-μm-i.d. nano-ESI tip (PicoTip, New Objective, Woburn, MA) at a flow rate of 200−250 nL/min. A total of 10−15 μL of sample was loaded by "pick-up" injection from the 96-well receiving tray of the digester which was placed in the HPLC autosampler (FAMOS). The tray was covered with a self-sealing silicone compression mat (Axygen Scientific, Union City, CA) and cooled to 10 °C to prevent sample evaporation. After injection, the sample slug was loaded onto a C18 "trap" column (5 μm, 300 Å, 300 μm i.d. × 1 mm, Nano-Precolumn, LC Packings/Dionex) using a loading pump (SWITCHOS) at a flow rate of 20 μL/min. The loading pump eluent was 0.5% acetic acid, 0.02% heptafluorobutyric acid in HPLC grade water. After 2 min of sample loading, a switching valve brought the trap in-line with the eluent of the HPLC pump (Ultimate), which had a flow rate was 0.245 mL/min with a 1000-

Table 1. Selected *Campylobacter* Strains Whose *hup* Gene Was Successfully Sequenced with the Exception of Those Strains That Do Not Have a *hup* Accession Number[a]

strain	synonym	species	source	location	HU type	*hup* accession no.
RM1221	ATCC BBA-1062	*C. jejuni*	chicken	USA (CA)	I	AY751947
RM1859		*C. jejuni*	unknown	unknown	II	AY751958
RM2228	ATCC BAA-1061	*C. coli*	chicken	USA	I	AY751981
RM1051	ATCC 43479	*C. coli*	human	Canada	II	AY751942
RM1891		*C. coli*	chicken	USA (TX)	III	AY751966
RM1865		*C. coli*	human	USA	IV	AY751961
RM2092		*C. upsaliensis*	human	unknown	I	AY751974
RM3776		*C. upsaliensis*	human	South Africa	II	AY752011
RM3195	ATCC BBA-1059	*C. upsaliensis*	human	South Africa	III	AY751993
RM4245	LMG 9108	*C. upsaliensis*	human	Belgium	IV	AY752070
RM2100	ATCC BBA-1060	*C. lari*	human	unknown	I	AY751978
RM2825	LMG 11760	*C. lari*	human	Canada	II	AY751990
RM1890	ATCC 43675	*C. lari*	human	unknown	II	
RM3807		*C. helveticus*	feline	USA (CA)		
RM3806		*C. concisus*	human	South Africa		

[a] All strains underwent protein biomarker extraction except *C. lari* strain RM2825.

to-1 split resulting in a flow through the trap, analytical column, and spray tip of 200−250 nL/min. Mobile phase A was 0.5% glacial acetic acid diluted in HPLC grade water. Mobile phase B was 80% HPLC grade acetonitrile, 20% HPLC grade water, 0.5% glacial acetic acid. The HPLC gradient was as follows: 0−12 (92−20% A), 12−16 (20−92% A), 16−29 (92% A), and 29−29.1 min (92% A). The analytical column used was a C18 monomeric column (Vydac, W. R. Grace & Co.).

The reflectron-TOF of the instrument was externally calibrated as previously described. Data were acquired using the data-dependent scanning of the instrument software (Instrument Dependent Acquisition, IDA, Analyst QS). The sequence of acquisition steps were as follows: (1) survey scan of the mass range, m/z 400−1500; (2) identification of the m/z of the most abundant doubly or triply charged ion (above a threshold of 20 counts/s) in the survey scan; (3) adjustment of the quadrupole mass filter (Q_1) to the selected m/z; (4) collision-activated dissociation of the mass-selected ions in the collision quadrupole (Q_2) with UHP nitrogen; (5) analysis of the fragment ions by the TOF mass analyzer over a mass range of m/z 70−2000. After product ion analysis, the m/z of the ion was placed on an exclude list for 45 s, during which time it was not selected for further MS/MS analysis. The process was then repeated with another survey scan identifying the most abundant doubly or triply charged ion (not on the exclude list) followed by MS/MS analysis. An IDA script was used to determine the optimum collision energy for each product ion experiment based on its m/z and z. The instrument parameters were as follows: accumulation time for survey scan, 1.0 s; accumulation time for product ion scan, 3.0 s; current gas, 20 arbitrary units (au); declustering potential, 50 au; focusing potential, 205 V; second declustering potential, 15 au; ion release delay, 11 au; ion release width, 10 au; collision gas, 9 au.

Database Searches. The WIFF acquisition files created by Analyst QS were converted to DTA files using a WIFF-to-DTA converter (Genomic Solutions, Ann Arbor, MI). The DTA files containing MS/MS data were then searched against a flat file containing amino acid sequences of selected HU proteins encoded from *hup* sequencing and amino acid sequences of all proteins encoded from genomic sequencing of *C. jejuni* strain RM1221, *C. jejuni* NCTC 11168 (NCBI), and *C. coli* strain RM2228. The software used was from The Global Proteome Machine organization (http://www.thegpm.org/). A local installation of that open-source project (GPM) was used for analysis of the data. The spectrum modeler X! TANDEM, which is a part of the GPM, was used to match MS/MS peptide fragmentation data to peptide sequences.[30]

RESULTS AND DISCUSSION

The *hup* genes of 143 *Campylobacter* strains were sequenced: 38 *C. jejuni*, 22 *C. coli*, 9 *C. lari*, and 74 *C. upsaliensis*. From these 143 *hup* genes, 12 unique amino acid sequences were found for HU. Two unique amino acid sequences were identified for HU for *C. jejuni*, four for *C. coli*, two for *C. lari*, and four for *C. upsaliensis*. It should be noted that tentative identification and classification of HU protein types among various species and strains of *Campylobacter* was initially postulated from MALDI-TOF-MS protein biomarker analysis[27,28] prior to extensive *hup* gene sequencing. Sequencing confirmed what was already suspected from the MALDI-TOF-MS analysis, and subsequent proteomic analysis confirmed both the gene sequencing and the MALDI-TOF-MS analysis. All 143 *hup* sequences were deposited into GenBank. Table 1 lists selected *Campylobacter* strains whose *hup* gene were successfully sequenced (exceptions were *C. lari* strain RM1890, *C. helveticus* strain RM3807, and *C. concisus* strain RM3806). The strains in Table 1 also underwent protein biomarker extraction (the only exception was *C. lari* strain RM2825).

Figure 1 (top panel) shows a typical MALDI-TOF-MS spectrum of *C. jejuni* strain RM1221 cell lysate. One of the most prominent ions in the spectra is at m/z 10 277. The ion appears to be singly charged, i.e., $[M + H]^+$, as its doubly charged counterpart also appears at m/z 5139, i.e., $[M + 2H]^{2+}$. A number of other biomarker ions are also present. The mass spectrum was externally calibrated. Figure 1 (middle panel) shows the MALDI-TOF-MS spectrum of HPLC fraction 75, which was collected from the cell lysate solution of *C. jejuni* strain RM1221. The spectrum

(30) Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20*, 1466−1467.

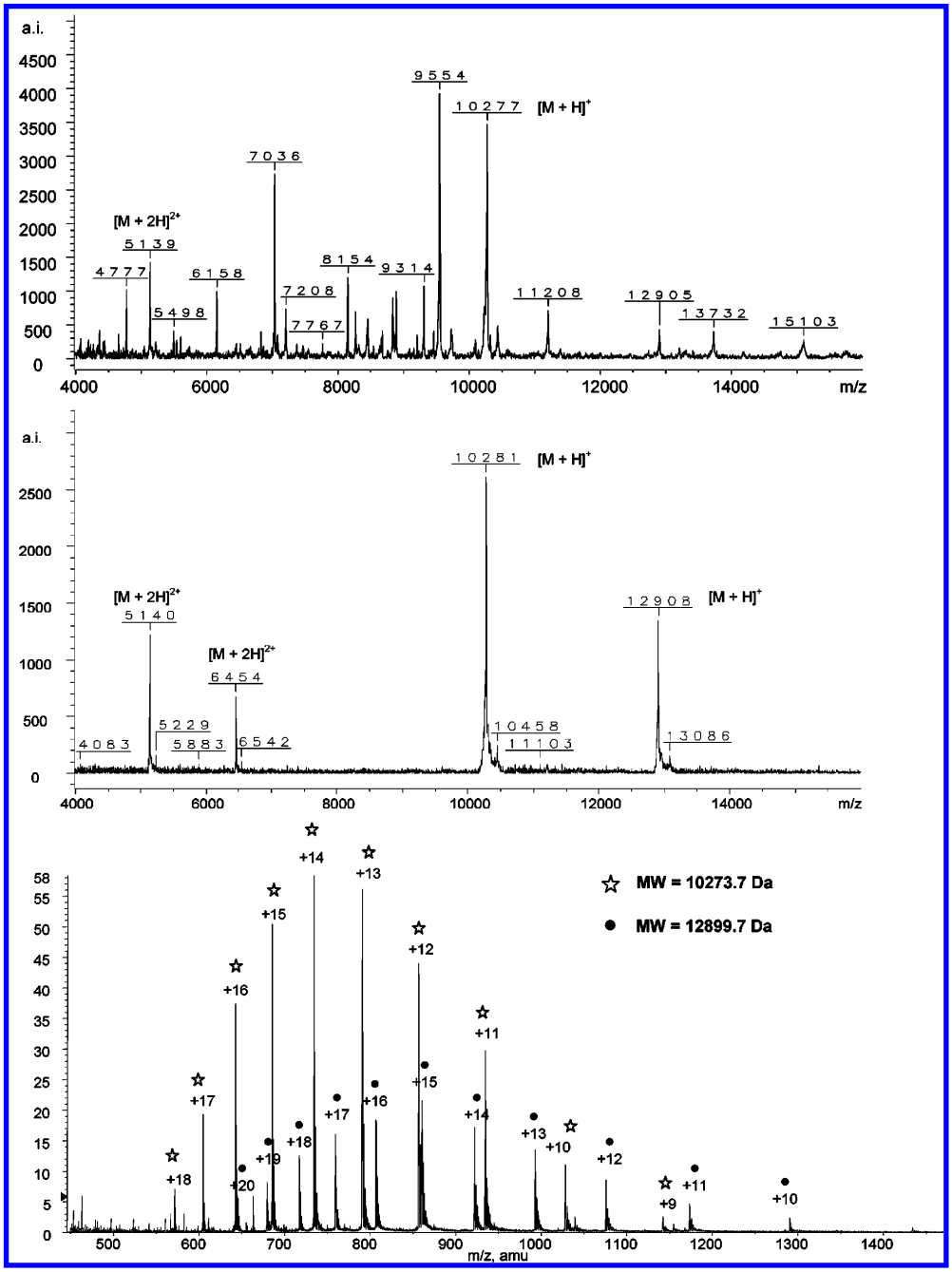

Figure 1. Top panel: MALDI-TOF-MS spectrum of *C. jejuni* strain RM1221 cell lysate. Middle panel: MALDI-TOF-MS spectrum of HPLC fraction 75 of *C. jejuni* strain RM1221 cell lysate. Bottom panel: high-resolution ESI-MS of HPLC fraction 75 of *C. jejuni* strain RM1221 cell lysate.

was again externally calibrated. As shown in the figure, two proteins were present in this HPLC fraction and both show a $[M + H]^+$ and $[M + 2H]^{2+}$ charge states. The ions at m/z 10 277 (top panel) and 10 281 (middle panel) were within the mass accuracy of our externally calibrated instrument, and their relative ion signals were also comparable. We concluded that the biomarker ion at m/z 10 277 is the same biomarker ion at m/z 10 281 as no other HPLC fraction contained an ion at this m/z. Figure 1 (bottom panel) shows the HR-ESI-MS analysis of the HPLC fraction 75. In this figure, two overlapping charge-state envelopes are observed. Deconvolution of the more abundant charge state envelope (marked by stars) gave a MW = 10 273.7, which is close to both the MALDI-TOF MW of the HPLC fraction (10 280) and the MALDI-TOF MW of the cell lysate (10 276). Deconvolution of the less abundant charge state envelope (marked by ●) gave a MW of 12 899.7, which is close to both the MALDI-TOF MW of the HPLC fraction (12 907) and the MALDI-TOF MW of the cell lysate (12 904).

Figure 2 shows a 1-D SDS PAGE gel of HPLC fraction 75 from *C. jejuni* strain RM1221 cell lysate (left lane). The most intense band is centered at ~10 kDa as estimated from the protein standards (right lane). The second most intense gel band is centered at ~13−14 kDa. A number of minor bands are also observed. All the major and minor bands in the lane were excised and digested with trypsin. The tryptic peptides of each band were then analyzed by nano-LC/MS/MS. When the MS/MS spectra

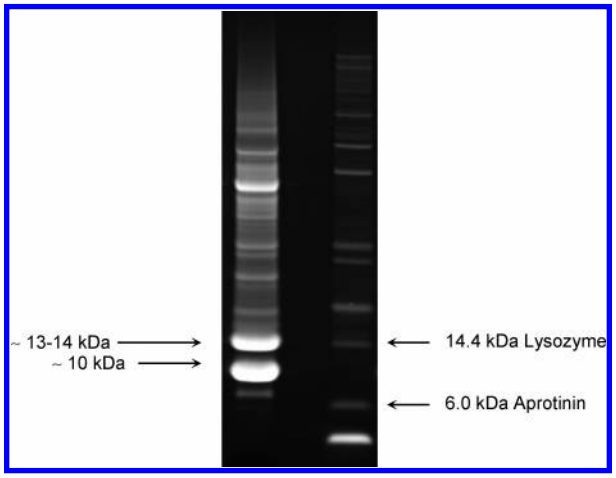

Figure 2. Left lane: 1-D SDS PAGE of HPLC fraction 75 of *C. jejuni* strain RM1221 cell lysate. Right lane: protein standards.

were compared against our in-house database of bacterial protein sequences (including HU amino acid sequences) using GPM software, the ~10-kDa gel band was identified as the HU protein of *C. jejuni*, strain RM1221 with a predicted MW of 10 274.0. No other excised gel band resulted in a protein identification with a MW of ~10 000. The second most intense gel band centered at ~13−14 kDa was identified as the 50S ribosomal protein L7/L12 with a predicted MW (after N-terminal methionine cleavage) of 12 899.8. The HR ESI-MS and nano-LC/MS/MS data on this protein will be reported in a subsequent publication.

Table 2 summarizes the results of the GPM analysis of the RM1221 ~10-kDa gel band with the probability scores of homologous HU proteins in the database identified by their *hup* accession number. There are, in fact, three strains of *C. jejuni* that are ranked with the highest score: RM1221, RM1859, and NCTC11168. NCTC 11168 and RM1221 have identical HU amino acid sequences although their nucleotide sequences are slightly different. The HU amino acid sequence for RM1859 is different from RM1221 by one residue. Table 3 summarizes the predicted amino acid sequence of each HU protein (in black and gray). The portion of the amino acid sequence that was confirmed by MS/MS is shown in black. We obtained 81.6% MS/MS sequence coverage of the RM1221 HU. Although this was sufficient coverage to distinguish the RM1221 HU from other *Campylobacter* HU proteins, it was not sufficient to distinguish it from the sequence of RM1859. The HU sequence for RM1221 and RM1859 are nearly identical except for a variation at residue 94. RM1221 has a glycine residue at this position; RM1859 has a serine. A single-point mutation in the nucleotide sequence is responsible for the residue variation. Table 4 lists the tryptic peptides used for MS/MS confirmation of the RM1221 HU amino acid sequence. The absence of a peptide that includes residues 92−98 results in our inability to distinguish between RM1221 and RM1859 by MS/MS alone. However, the 30-Da mass difference between glycine and serine should be reflected in a mass difference in the MWs of HU for RM1221 and RM1859. This was found to be the case as shown in Table 5. Table 5 also shows the MWs of HU predicted from the *hup* sequences. The measured and predicted HU MWs for RM1221 are in excellent agreement. Correspondingly, MS/MS analysis of the ~10-kDa gel band of *C. jejuni* RM1859 resulted in the identification of three HU strains with equally high ranking scores: RM1859, RM1221, and NCTC 11168 (all *C. jejuni*) as shown in Table 2. Although the scores were higher for the analysis of RM1859 than for the analysis of RM1221, the 72.4% MS/MS coverage (Table 3) was still not sufficient to distinguish it from RM1221 although it was easily distinguished from other *Campylobacter* species. The 30-Da mass difference of the glycine/serine variation for the 94th residue position resulted in a 30-Da mass difference in the HU MWs of RM1859 and RM1221, thus distinguishing these two biomarker ions.

Four tentative HU types were identified for *C. coli* from MALDI-TOF-MS protein biomarker analysis (Table 5), which were confirmed from amplification and sequencing of 22 *C. coli hup* genes. One strain from each type was selected for protein biomarker extraction: RM2228 (type I), RM1051 (type II), RM1891 (type III), and RM1865 (type IV). As shown in Table 2, when each of the extracted/digested ~10-kDa proteins of types I−IV were analyzed by nano-LC/MS/MS and compared against our in-house database, only one HU sequence (the strain processed for protein biomarker extraction) had the highest probability score. Thus, it was possible to differentiate the four HU sequences by MS/MS confirmation alone. Single-point mutations account for intraspecies variations at residues 22, 37, and 90 as shown in Table 3. Variation at residue 53 was caused by a two-base pair (bp) change. The MS/MS coverage (%) was as low as 65.6% for RM1891 to as high as 96.9% for RM1051. Intraspecies residue variations also resulted in unique HU MWs, which were

Table 2. Identification of the Homologues of the DNA-Binding Protein HU of *Campylobacter* from MS/MS Data Searched against an In-House Database[a]

C. jejuni	*C. coli*	*C. upsaliensis*	*C. lari*	*C. helveticus*
RM1221 -160.2 AY751947, AY751958, I40616	**RM2228** -172.1 AY751981	**RM2092** -91.5 AY751974, AY752070, AY752011	**RM2100** -159.1 AY751978	**RM3807** -53.4 AY751974, AY752070, AY752011
RM1859 -214.5 AY751947, AY751958, I40616	**RM1051** -280.9 AY751942	**RM3776** -257.4 AY752011	**RM1890** -243.4 AY751990	
	RM1891 -181.3 AY751966	**RM3195** -155.9 AY751993		
	RM1865 -160.9 AY751961	**RM4245** -264.8 AY752070		

[a] Highest probability score (log(e)) is reported along with the identified *hup* accession number(s).

Table 3. Alignment Table of the Predicted Amino Acid Sequence of HU (as Encoded by *hup* Gene) for Selected Species and Strains of *Campylobacter*[a]

Genus & Species	Strain	Sequence: 1-30	
C. jejuni	RM1221	**MTKADFISLVAQTAGLTKKDATTATDAVIS**	
"	RM1859	**MTKADFISLVAQTAGLTKKDATTATDAVIS**	
C. coli	RM2228	**MTKADFISQVAQTAGLTKKDATAATDAVIS**	
"	RM1051	**MTKADFISQVAQTAGLTKKDATAATDAVIS**	
"	RM1891	**MTKADFISQVAQTAGLTKKDAAAATDAVIS**	
"	RM1865	**MTKADFISQVAQTAGLTKKDATAATDAVIS**	
C. upsaliensis	RM2092	MTK**ADFISQVAQNAGLTKKDAGAATDAVIS**	
"	RM3776	**MTKADFISQVAQNAGLTKKDAGAATDAVIS**	
"	RM3195	**MTKADFISQVAQNAELTKKDAGAATDAVIS**	
"	RM4245	**MTKADFISQVAQNAGLTKKDAGAATDAVIS**	
C. lari	RM2100	**MTKADFISQVAQTAGLTKKDAGAATDAVIA**	
"	RM1890	**MTKADFISQVAQTAGLTKKDATAATDAVIA**	
		31-60	
C. jejuni	RM1221	**TITDVLAKGDSISFIGFGTFSTQERAAR**EA	
"	RM1859	**TITDVLAKGDSISFIGFGTFSTQER**AAREA	
C. coli	RM2228	**TITDVLAKGDSVSFIGFGTFSTAER**AAREA	
"	RM1051	**TITDVLTKGDSVSFIGFGTFSTAER**AAR**EA**	
"	RM1891	**TITDVLAKGDSVSFIGFGTFSTAER**AAREA	
"	RM1865	**TITDVLAKGDSVSFIGFGTFSTTER**AAREA	
C. upsaliensis	RM2092	**TITEVLAKGDSISFIGFGTFSTTER**AAREA	
"	RM3776	**TITEVLAKGDSISFIGFGTFSTTER**AAR**EA**	
"	RM3195	**TITEVLAKGDSISFIGFGTFSTTER**AAR**EA**	
"	RM4245	**TITEVLAKGDSISFIGFGTFSTTER**AAR**EA**	
C. lari	RM2100	**TITDVLAKGDSISFIGFGTFSVAER**AAR**EA**	
"	RM1890	**TITDVLAKGDSISFIGFGTFSVAER**AAR**EA**	
		61-90	
C. jejuni	RM1221	R**VPSTGKTIKVPATR**VAKFKVGK**NLKE**AVA	
"	RM1859	RVPSTGK**TIKVPATR**VAKFKVGK**NLKE**AVA	
C. coli	RM2228	R**VPSTGKTIKVPATRVAKFKVGKNLKDAVA**	
"	RM1051	**RVPSTGKTIKVPATRVAKFKVGKNLKDAVA**	
"	RM1891	RVPSTGKTIKVPATRVAKFKVGK**NLKDAVA**	
"	RM1865	R**VPSTGKTIKVPATR**VAK**FKVGKNLKDAV**V	
C. upsaliensis	RM2092	RVPSTGKTIKVPATRVAKFKVGK**S**LK**E**AVA	
"	RM3776	**RVPSTGKTIKVPATRVAKFKVGKNLKE**AVA	
"	RM3195	**RVPSTGKTIKVPATRVAKFKVGKNLKE**AVA	
"	RM4245	**RVPSTGKTIKVPATRV**V**KFKVGKNLKE**AVA	
C. lari	RM2100	**RVPSTGA**TIKVPAT**K**VAKFKVGK**NLKDAVA**	
"	RM1890	**RVPSTGA**TIKVPAT**K**VAKFKVGK**NLKDAVA**	
		91-96, 98, 100	MS/MS Coverage (%)
C. jejuni	RM1221	**—KA**S**GKKKK**	81.6
"	RM1859	**—KA**S**S**KKKK	72.4
C. coli	RM2228	**—KA——KKK**	88.5
"	RM1051	**—KA——KKKK**	96.9
"	RM1891	**—KA——KKK**	65.6
"	RM1865	**—KA——KKK**	85.4
C. upsaliensis	RM2092	A**KA—GKKKK**	53.1
"	RM3776	**AKA—GKKKK**	93.9
"	RM3195	**AKA—GKKKK**	93.9
"	RM4245	**AKA—GKKKK**	94.9
C. lari	RM2100	**AAK**T**AKKAKK**	90.0
"	RM1890	**AAK**T**AKKAKK**	90.0

[a] Black boldface type indicates sequence was confirmed by nano-LC-ESI/MS/MS. The MS/MS confirmation for RM1890 was compared to the predicted primary amino acid sequence of HU for RM2825 (another *C. lari* type II HU strain). Residues in the minority are boxed.

confirmed by the measured and predicted MWs shown in Table 5.

Four tentative HU types were identified for *C. upsaliensis* from MALDI-TOF-MS protein biomarker analysis (Table 5) and were later confirmed by amplification and sequencing of 74 *C. upsaliensis hup* genes. One strain from each type was selected for protein biomarker extraction: RM2092 (type I), RM3776 (type II), RM3195 (type III), and RM4245 (type IV). Like *C. coli*, it was possible to differentiate by MS/MS confirmation alone that part of the HU sequence that was unique to each *C. upsaliensis* strain (except for RM2092), which was reflected in a highest probability score shown in Table 2. As shown in Table 3, the MS/MS confirmation coverage of each strain was over 93% (except for RM2092, which had 53.1% coverage). Single-point mutations account for variations in the HU amino acid sequence of the four strains at residues 15, 77, and 84, and these variations resulted in unique HU MWs for each strain as shown in Table 5. It should be noted that the HU for *C. upsaliensis* strain RM3195 is near enough in MW to the HU for *C. jejuni* strain RM1221 as to be indistinguishable when analyzed on our MALDI-TOF-MS by external calibration. Internal calibration with its increased mass accuracy (±1 Da) is useful for distinguishing biomarkers with

Table 4. List of Tryptic Peptides from a ~ 10-kDa Protein Biomarker of *C. jejuni* Strain RM1221 Analyzed by Nano-LC/MS/MS and Used To Identify the Biomarker as HU

Sequence	Mass	Charge	# of MS/MS on Peptide	Highest Score Log (e)
[1]**M**TKADFISLVAQTAGLTKK[19]	2039.1	+2	1	-11.0
[1]**M**TKADFISLVAQTAGLTKK[19]	2039.1	+3	3	-6.6
[1]**M**TKADFISLVAQTAGLTK[18]	1911.0	+2	1	-5.2
[1]**M**TKADFISLVAQTAGLTK[18]	1911.0	+3	1	-4.5
[4]ADFISLVAQTAGLTKK[19]	1662.9	+2	4	-6.8
[4]ADFISLVAQTAGLTKK[19]	1662.9	+3	10	-4.9
[4]ADFISLVAQTAGLTK[18]	1534.8	+2	9	-6.6
[4]ADFISLVAQTAGLTK[18]	1534.8	+3	9	-10.0
[6]FISLVAQTAGLTKK[19]	1476.9	+2	1	-3.3
[6]FISLVAQTAGLTKK[19]	1476.9	+3	1	-1.0
[6]FISLVAQTAGLTK[18]	1348.8	+2	1	-3.6
[8]SLVAQTAGLTKK[19]	1216.7	+2	1	-3.3
[18]KKDATTATDAVISTITDVLAK[38]	2204.2	+3	1	-9.2
[19]KDATTATDAVISTITDVLAK[38]	2034.1	+2	1	-8.8
[19]KDATTATDAVISTITDVLAK[38]	2034.1	+3	6	-8.8
[20]DATTATDAVISTITDVLAK[38]	1906.0	+2	7	-10.5
[20]DATTATDAVISTITDVLAK[38]	1906.0	+3	7	-8.0
[20]DATTATDAVISTITD[34]	1494.7	+2	1	-1.5
[21]ATTATDAVISTITDVLAK[38]	1791.0	+3	1	-6.1
[27]AVISTITDVLAK[38]	1230.7	+2	1	-5.1
[27]**A**VISTITDVLAK[38]	1272.7	+2	1	-2.8
[39]GDSISFIGFGTFSTQERAAR[58]	2147.1	+3	1	-2.0
[39]GDSISFIGFGTFSTQER[55]	1848.9	+2	9	-6.4
[39]GDSISFIGFGTFSTQER[55]	1848.9	+3	11	-6.1
[39]GDSISFIGFGTF[50]	1247.6	+2	1	-6.2
[45]IGFGTFSTQER[55]	1242.6	+2	1	-3.9
[62]VPSTGKTIKVPATR[75]	1454.9	+3	1	-1.3
[84]NLKEAVAK[91]	872.5	+2	1	-1.4

very close MWs (although multiple biomarkers are used for pathogen identification purposes).[27,28]

Two tentative HU types were identified for *C. lari* from MALDI-TOF-MS protein biomarker analysis (Table 5), which were later confirmed by amplification and sequencing of 9 *C. lari hup* genes. One strain from each HU type was selected for protein biomarker extraction: RM2100 (type I) and RM1890 (type II). It should be noted that attempts to amplify and sequence RM1890 *hup* were not successful due to differences in the flanking genes in that strain. Consequently, identification of RM1890 as a *C. lari* HU type II was based on MS and MS/MS analysis and comparison to another *C. lari* HU type II encoded by the RM2825 *hup* gene. As shown in Table 2, *C. lari* HU types I and II are differentiated by MS/MS alone, and the highest probability score for RM1890 strain is the RM2825 sequence. MS/MS confirmation coverage was 90% for both strains, which included a glycine/threonine (G/T) residue variation at position 22 caused by a two nucleotide basepair change. That single variation resulted in a HU MW difference between the two *C. lari* HU types of 44 Da (Table 5).

We also extracted the protein biomarker for *C. helveticus* strain RM3807 even though we were unsuccessful in sequencing its *hup* gene. However, MS/MS analysis of its protein biomarker showed significant conservation in the HU sequence of *C. upsaliensis* strains RM2092, RM4245, RM3776, and RM3195, which was reflected in the probability scores in Table 2 and the MS/MS confirmation coverage shown in Table 6. *C. helveticus* and *C. upsaliensis* are known to be phylogenetically close; thus, finding conserved HU sequence between these two species was not surprising. However, multilocus sequence typing definitively proved that RM3807 is a *C. helveticus*.[29]

We also extracted a ~10.5-kDa protein biomarker from *C. concisus* strain RM3806. Attempts at sequencing the *hup* gene of RM3806 were also unsuccessful. However, in contrast with *C. helveticus*, MS/MS analysis of the extracted/digested ~10.5-kDa protein showed no significant protein identification when compared against our database. As *C. consisus* is phylogenetically distant from the thermophilic *Campylobacters*, this was not unexpected. Attempts are now underway in our laboratory to de novo sequence this tentatively assigned HU protein biomarker from MS/MS data alone.

Genomic and Proteomic Identification of Protein Biomarkers. The strength of combining both genomics and proteomics for identification of a MALDI-TOF-MS protein biomarker is best highlighted by examining the limitations on the use of either technique in isolation. First, the existence of a gene is not proof that the protein is expressed, or expressed in sufficient numbers such that it will result in a detectable ion signal by MALDI-TOF-MS. Second, although it is possible to tentatively assign a biomarker ion on the basis of a comparison to a database, it is not proof that the assignment is correct. Two or more proteins within a genome may have identical (or very close) MWs, and this problem becomes even more significant when comparing the protein MWs of multiple genomes. In that regard, MALDI-TOF mass accuracy plays a critical role in the confidence of such an assignment process. Third, PTMs may occur which can cause a protein biomarker's actual MW to be different from its genomically predicted MW. In consequence, the bioinformatics approach of pathogen identification, which relies exclusively on a comparison of observed and hypothetical protein MWs, requires assignment of multiple protein biomarkers. Algorithms used in this approach have incorporated PTMs that are common to microbial proteins, e.g. N-terminal methionine cleavage.[21]

By the same token, identification of a MALDI-TOF protein biomarker by protein de novo sequencing alone (i.e., no database searching) is challenging, labor-intensive, and prone to possible error. De novo sequencing involves MS/MS of the tryptic peptides of the protein such that one obtains (if possible) full sequence coverage of the entire protein with high confidence of the sequence of many if not all the peptides analyzed. It also requires identification of PTMs (if present). By this approach, the amino acid composition of the final de novo protein sequence (with PTMs) must concur exactly with the observed protein MW. Confirmation of the protein biomarker function can then be tentatively determined from a BLAST search.

HU as a Protein Biomarker. HU is particularly useful as a biomarker for a number of reasons. First, it has been a prominent biomarker ion in hundreds of MALDI-TOF-MS spectra collected from cell lysates of various species and strains of *Campylobacter* in our laboratory over a number of years.[27] The prominence of HU in MALDI-TOF spectra of cell lysates is likely due to its high copy number as well as the four lysine residues clustered at the carboxyl end of the protein. These four lysine residues likely ensure high ionization efficiency of the protein by MALDI. Second, the HU amino acid sequence varies between species (and sometimes between strains of the same species), which is a direct result of variations in the *hup* sequences. The amount of HU sequence variation observed between species is reflected in their apparent phylogenetic proximity. For instance, *C. helveticus* is

Table 5. MW of the ~10-kDa Protein Biomarker As Measured by MALDI-TOF-MS of Cell Lysate and by High-Resolution ESI-MS from an HPLC Fraction

genus and species	strain	HU type	MALDI-TOF MS MW	HR ESI-MS MW	predicted[a] HU MW
C. jejuni	RM1221	I	10 276	10 273.9 ± 0.2 ($n = 3$)	10 274.0
C. jejuni	RM1859	II	10 304	10 304.0 ± 0.2 ($n = 4$)	10 304.0
C. coli	RM2228	I	10 031	10 029.4 ± 0.2 ($n = 3$)	10 029.7
C. coli	RM1051	II	10 059	10 059.5 ± 0.1 ($n = 3$)	10059.7
C. coli	RM1891	III	10 004	9 999.7 ± 0.1 ($n = 3$)	9 999.6
C. coli	RM1865	IV	10 090	10 087.7 ± 0.3 ($n = 3$)	10 087.7
C. upsaliensis	RM2092	I	10 170	10 171.7 ± 0.1 ($n = 4$)	10 171.8
C. upsaliensis	RM3776	II	10 196	10 198.7 ± 0.1 ($n = 3$)	10 198.9
C. upsaliensis	RM3195	III	10 271	10 270.8 ± 0.3 ($n = 3$)	10 270.9
C. upsaliensis	RM4245	IV	10 228	10 226.8 ± 0.2 ($n = 3$)	10 226.9
C. lari	RM2100	I	10 217	10 210.8 ± 0.1 ($n = 3$)	10 210.9
C. lari	RM1890	II	10 254	10 255.0 ± 0.3 ($n = 4$)	10 255.0
C. helveticus	RM3807		10 190	10 186.9 ± 0.3 ($n = 3$)	
C. concisus	RM3806		10 519	10 516.3 ± 0.1 ($n = 3$)	

[a] The predicted MW of HU as determined by *hup* sequencing and calculated using GPMAW software (Version 5.10, Lighthouse Data).

Table 6. MS/MS Coverage (Top Panel)[a] and a List of Tryptic Peptides from a ~10-kDa Protein Biomarker of *C. Helveticus* Strain RM3807 (Bottom Panel)[b]

Genus & Species	Strain	Sequence: 1-30	
C. helveticus	RM3807	-- -- -- -- -- -- -- -- **Q V A Q N A G L T K K D A G A A T D A V I S**	
		31-60	
"	"	**T I T E V L A K G D S I S F I G F G T F S T T E R** -- -- -- -- --	
		61-90	
"	"	-- --	
		91-98	MS/MS Coverage (%)
"	"	-- -- -- -- -- -- -- --	~ 50

Sequence	Mass	Charge	# of MS/MS on Peptide	Highest Score Log (e)
9**Q**VAQNAGLTKKDAGAATDAVISTITEVLAK38	3026.6	3	1	-2.6
19**K**DAGAATDAVISTITEVLAK38	2016.1	3	2	-2.3
20**D**AGAATDAVISTITEVLAK38	1888.0	2	9	-8.9
"	"	3	13	-3.7
21**A**GAATDAVISTITEVLAK38	1773.0	3	1	-2.4
27AVISTITEVLAK38	1244.7	2	1	-4.9
39GDSISFIGFGTFSTTER55	1821.9	2	18	-11.3
"	"	3	12	-7.0
41SISFIGFGTFSTTER55	1649.8	2	1	-5.9

[a] The amino acid sequence is partially conserved in the sequences of HU of *C. upsaliensis* strains. [b] Analyzed by nano LC/MS/MS and used to identify the biomarker as HU from partially conserved HU sequences of *C. upsaliensis* strains. Boldfaced letters are modified residues.

closer, phylogenetically, to *C. upsaliensis* than to the other thermophilic *Campylobacters*: *C. jejuni*, *C. coli*, or *C. lari*. As shown in Table 6, we found partial conservation of the HU sequence for *C. helveticus* strain RM3807 in the HU sequences of *C. upsaliensis* strains RM2092, RM3776, RM3195, and RM4245 but not in the HU amino acid sequence of other thermophilic *Campylobacter* species. *C. concisus* (RM3806), which is phylogenetically distant from all thermophilic *Campylobacters*, produced no apparent conservation of HU sequence in the HU sequences (or other protein sequences) of other *Campylobacters* on the basis of MS/MS data. Third, the *Campylobacter* HUs, in this study, have no PTMs, not even the most common PTM of microbial proteins, i.e., N-terminal methionine cleavage. The absence of this PTM for HU is unusual in that the penultimate residue of HU, i.e.,

threonine, would be expected to facilitate posttranslational cleavage of methionine by bacterial aminopeptidases.[21,31−33] Presumably, the sequence (or structure) of HU may preclude its undergoing this PTM. Interestingly, Hathout et al.[34] reported posttranslational N-terminal methionine cleavage from small, acid-soluble, nonspecific DNA-binding proteins of *Bacillus* spores analyzed by MALDI-TOF-MS. The penultimate residues of these proteins (alanine, glycine, or serine) would also be expected to facilitate *N*-Met removal,[21,31−33] which is exactly what was reported. Although *N*-Met removal is not observed in *Campylobacter* HUs, the aminopeptidases responsible for *N*-Met removal appear to be present and functioning in *Campylobacter* as indicated by *N*-terminal methionine cleavage from 50 S ribosomal protein L7/L12 of strain RM1221 noted earlier. In addition, Mandrell, Harden, and co-workers reported detection of possible *N*-Met removal in other *Campylobacter* protein biomarkers by comparison to protein MWs in a genomically derived database.[27] In some cases, biomarker assignment from the database was only possible if one assumed a mass loss of 131 Da (the mass of methionine) from the protein. The biomarkers assigned in this way were often tentatively identified as ribosomal proteins. Definitive identification of these and other *Campylobacter* protein biomarkers will be reported in a subsequent communication.

(31) Hirel, P. H.; Schmitter, J. M.; Dessen, P.; Fayet, G.; Blanquet, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8247−8251.
(32) Gonzalez, T.; Baudouy, J. *J. FEMS Microbiol. Rev.* **1996**, *18*, 319−334.
(33) Solbiati, J.; Chapman-Smith, A.; Miller, J.; Miller, Ch.; Cronan, J., Jr. *J. Mol. Biol.* **1999**, *290*, 607−614.
(34) Hathout, Y.; Setlow, B.; Cabrera-Martinez, R.-M.; Fenselau, C.; Setlow, P. *Appl. Environ. Microbiol.* **2003**, *69*, 1100−1107.

CONCLUSIONS

MALDI-TOF-MS protein biomarker analysis of bacterial cell lysates provides a relatively simple and rapid method for identifying and discriminating *Campylobacter* species and strains.[27,28] The current work demonstrates that it is possible to isolate prominent protein biomarkers observed in the MALDI-TOF-MS spectra and unambiguously identify them by a combination of genomic and proteomic analysis. Based upon our analysis, we have found that a DNA-binding protein HU is extremely useful as a biomarker for discriminating *Campylobacter* species and strains due to variability of its primary amino acid sequence among species and strains. In contrast with other bacterial proteins, HU does not undergo any posttranslational modifications, including cleavage of the N-terminal methionine. The cluster of four lysine residues at the C-terminal end of the protein ensures high ionization efficiency by "soft" ionization techniques.

ACKNOWLEDGMENT

We thank Dr. William F. Haddon for his prior contributions to the early stages of this project. We also thank Dr. Bruce Onisko and Dr. Chris Silva for useful discussions. Mention of a brand or firm name does not constitute an endorsement by the U.S. Department of Agriculture over other of a similar nature not mentioned.

Received for review December 28, 2004. Accepted May 18, 2005.

AC040193Z