Proteomic and genomic analysis reveals novel Campylobacter jejuni outer membrane proteins and potential heterogeneity

Eleanor Watsona,∗, Aileen Sherryb, Neil F. Inglisa, Alex Lainsona, Dushyanth Jyothia,1, Raja Yagaa, Erin Mansona, Lisa Imriea,2, Paul Everestb, David G.E. Smitha,b

a Moredun Research Institute, Bush Loan, Penicuik, United Kingdom
b Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

Article info
Article history:
Received 9 May 2014
Accepted 19 June 2014
Available online 26 June 2014

Keywords:
Campylobacter jejuni
Proteomics
Outer membrane
Genomics

Abstract
Gram-negative bacterial outer membrane proteins play important roles in the interaction of bacteria with their environment including nutrient acquisition, adhesion and invasion, and antibiotic resistance. In this study we identified 47 proteins within the Sarkosyl-insoluble fraction of Campylobacter jejuni 81-176, using LC–ESI-MS/MS. Comparative analysis of outer membrane protein sequences was visualised to reveal protein distribution within a panel of Campylobacter spp., identifying several C. jejuni-specific proteins. Smith–Waterman analyses of C. jejuni homologues revealed high sequence conservation amongst a number of hypothetical proteins, sequence heterogeneity of other proteins and several proteins which are absent in a proportion of strains.

© 2014 The Authors. Published by Elsevier B.V. on behalf of European Proteomics Association (EuPA). This is an open access article under the CC BY license.

1. Introduction
Campylobacter spp. are the commonest cause of food-borne disease worldwide accounting for 2.4 million cases per year in the US (www.cdc.gov/foodnet) with Campylobacter jejuni and C. coli responsible for the majority of infections. In the developing world these organisms are the leading cause of gastrointestinal infection in children under 2 years of age [1]. Additionally, post-infection sequelae may arise including Guillain–Barré syndrome (GBS) and other debilitating neurological disorders [2,3]. Despite a significant worldwide effort, mechanisms of disease and immunity remain poorly understood. Gram-negative bacterial outer membrane (OM) proteins represent a group of factors which play important roles in the interaction of bacteria with their environment. These include porins and nutrient uptake systems, iron acquisition proteins,
virulence factors, proteins involved in antibiotic resistance and other proteins required for survival within the hostile in vivo environment.

To date, functional roles have been ascribed to only a dozen or so C. jejuni OM proteins. A more comprehensive characterisation of this important group of proteins, even in prototypic strains, has yet to be reported. Several C. jejuni surface proteins have been observed to play roles in adhesion. Three of these, CadF and Flpa, and PEB1 have been shown to play a role in the colonisation of broiler chicks [4,5]. Additionally, JlpA, CapA and PorA appear to mediate binding to epithelial cells in culture [6–8]. Another important role is the transportation of small molecules across the bacterial cell membrane. C. jejuni lacks the phosphofructokinase protein and cannot therefore metabolise glucose although recently some strains were shown to use fucose as a substrate for growth [9,10]. Instead, the bacterium utilises amino acids as a source of carbon and energy; the surface proteins PEB1, and CjaA are components of ABC transporters with roles in aspartate/glutamate, and cysteine uptake respectively [11,12]. OM efflux systems are also important: CmeC and CmeD are OM components of two functionally characterised multi-drug efflux pumps, which have roles in antimicrobial and bile resistance [13,14]. Omp50 and PorA have been characterised as porins and the former was recently shown to play a role in a phosphotyrosine network regulation [15–17]. Campylobacter spp. also possess a number of iron acquisition systems, components of which reside in the OM including the Fe3+ enterobactin binding proteins CfrA and CfrB [18–20] and the heme transporter ChuA, [21]. Flagella have a variety of proposed roles in C. jejuni virulence and protein components which anchor flagella to the membrane (FlgH) are located in the OM [22]. OM proteins may also elicit pathological host responses; the surface-exposed lipoprotein JlpA is believed to trigger signalling events, which lead to inflammation [23].

Interrogation of the C. jejuni genome has revealed that a number of protein complexes and chaperones, which are essential for membrane biogenesis in many well characterised Gram-negative bacteria, are either divergent or absent altogether within C. jejuni genomes [24]. The β-barrel assembly machinery is known as the Bam complex [25]. The closest homologue of the OM localised component, BamA (formerly known as YaeT), is divergent in C. jejuni and, to the best of our knowledge, homologues of BamBCDE, are not found at all within Epsilonproteobacteria genomes. LoLB of the Lol (lipoprotein localisation) complex, which is responsible for sorting lipoproteins [26] is absent in Epsilonproteobacteria, although homologues of other proteins within this complex are present in C. jejuni genomes. C. jejuni also varies from the majority of other Gram-negative bacteria in that it lacks the O-antigen of LPS. Often referred to as LOS (lipooligosaccharide), this is perhaps a reflection of the divergence of the Lpt (LPS transport) proteins which are required for insertion of LPS into the outer membrane. This suggests that alternative mechanisms of maintaining membrane integrity and biogenesis have yet to be revealed within C. jejuni and related bacteria. Existing evidence suggests that distinct mechanisms for OM biogenesis exist outside the Gammaproteobacteria group of bacteria [27,28]. The divergence in this cellular machinery for protein sorting also suggests that conventional in silico protein localisation tools may not be reliable for this particular group of bacteria and therefore localisation of these proteins by experimental methods is particularly pertinent.

Approximately one third of C. jejuni predicted open reading frames (ORFs) code for proteins of unknown function and it is unlikely that the current list of characterised OM proteins is complete. Therefore, a thorough analysis of expressed OM proteins is essential for identifying factors important to and possibly novel to C. jejuni pathogenesis. We performed proteomic analysis of the OM of C. jejuni subsp. jejuni 81-176, a widely employed strain which causes experimentally reproducible clinical disease [29]. A variety of methods have been used previously for membrane protein enrichment although the fractions of extracted proteins vary in composition. The most comprehensive proteome analysis of C. jejuni to date focussed on the entire membrane compartment, i.e. periplasm and both inner and outer membranes [30]. Campylobacter OM appears to be closely associated with the inner membrane and as a result is more challenging to purify [31]. Recently, Hobbs et al. [32] reported that N-lauroylsarcosine (Sarkosyl) treatment of C. jejuni cells was the most successful method of isolating specifically OM proteins. Furthermore, Sarkosyl enrichment is often used to predict localisation of Campylobacter proteins [15,33,34].

The low solubility of OM proteins renders them incompatible with the majority of proteomic techniques hence gel-based proteomics offers a convenient method for their analysis as issues associated with hydrophobicity are circumvented by ionic detergent (sodium dodecyl sulfate) solubilisation and subsequent in-gel tryptic digestion [35,36]. In this study, a rapid shotgun proteomics-based approach was used to catalogue the protein complement of the C. jejuni OM fraction. This methodology, comprising SDS-PAGE, one-dimensional monolithic column liquid chromatography, electrospray ionisation (ESI) and fast MS/MS scanning, is colloquially termed “sawn-off shotgun proteomic analysis” (SOSPA). This approach enables the analysis of membrane-associated and other hydrophobic proteins whilst simultaneously combining rapidity with breadth of coverage. Bioinformatic approaches were deployed to survey the resulting SOSPA-generated data to identify homologous proteins amongst bacterial, Epsilonproteobacteria and, particularly, Campylobacter genomic sequences.

2. Materials and methods

2.1. Bacterial strains, media and culture conditions

C. jejuni 81-176 (pVir+) is a well characterised strain, isolated from contaminated milk [37]. Bacteria were grown at 37 °C in a variable atmosphere incubator (Don Whitley Scientific, Shipley, UK) in an atmosphere of 6% hydrogen, 5% carbon dioxide, 5% oxygen, and 84% nitrogen. Bacteria were cultured for 48 h on Campylobacter selective agar (Skirrow) plates (E&O Laboratories, Bonnybridge, UK) then resuspended in 20 ml high glucose Dulbecco’s modified Eagle’s medium (cat. 11960, Invitrogen, Paisley, UK), supplemented with 20 mM L-glutamic acid and 0.00125% iron ascorbate (Sigma, Dorset, UK) [38] at an Abs600 of
0.1. Cultures were incubated statically for 34 h; the timepoint was chosen to represent growth in late log phase.

2.2. OM protein enrichment

OM protein enrichment was carried out as described by Gauthier et al. [39]. Bacteria from broth cultures were pelleted at 8000 × g and resuspended in 500 μl of 50 mM Tris–HCl (pH 7), with 20% sucrose, 10 μM EDTA, 10 μg/ml lysosome and protease inhibitors (Complete Mini EDTA Free, Roche Diagnostics Ltd, Burgess Hill, UK) at 10 min at room temperature. All subsequent steps were carried out on ice. Bacteria were centrifuged at 8000 × g for 10 min and the supernatant removed. The pellet was resuspended in 1 ml Tris buffer (10 mM Tris–HCl, pH 7) with protease inhibitors and sonicated on ice (amplitude 5.0, 6 × 15 s). Unbroken bacteria were removed by centrifugation at 16,000 × g for 2 min and the supernatant was centrifuged for 1 h at 50,000 × g to pellet bacterial membranes. Pelleted membranes were washed with Tris buffer, resuspended in Tris–Sarkosyl buffer (Tris buffer containing 0.5% (w/v) N-lauroylsarcosine), and centrifuged for 1 h at 50,000 × g. The OM pellet was washed in Tris–Sarkosyl buffer then resuspended in Tris–Sarkosyl buffer containing 0.1% SDS.

2.3. LC–ESI-MS/MS

Sarkosyl-insoluble proteins, prepared as described above, were resolved on a 4–12% Bis–Tris NuPAGE SDS-PAGE gradient gel (Invitrogen, Paisley, UK) in MES Buffer at 200 V (constant voltage) over 45 min. Proteins were visualised using Colloidal Coomassie Blue G250 (Sigma, Dorset, UK). Molecular size standards were included routinely on gels. LC–ESI-MS/MS was carried out essentially as described by Batycka et al. [40]. For each sample lane, a series of gel slices of equal size (2.5 mm), covering the entire lane, were excised from the SDS-PAGE gel before performing standard in-gel de-staining, reduction, alkylation and trypsinolysis procedures [41]. The samples were transferred to HPLC sample vials and stored at +4 °C until required for LC–ESI-MS/MS analysis. Liquid chromatography was performed using a Dionex Ultimate 3000 nano-HPLC system (Thermo Fisher Scientific, Hemel Hempstead, UK) comprising a WPS-3000 well-plate micro auto sampler, a FLM-3600 dual-gradient micropump and an SRD-3600 solvent rack controlled by Chromelon chromatography software (www.thermoscientific.com/dionex). A micro-pump flow rate of 246 μl/min was used in combination with a cap-flow splitter cartridge, affording a 1/82 flow split and a final flow rate of 3 μl/min through a 5 cm × 200 μm ID monolithic reversed phase column (Thermo Fisher Scientific, Hemel Hempstead, UK) maintained at 50 °C. Samples of 4 μl were applied to the column by direct injection. Peptides were eluted by the application of a 15 min linear gradient from 8% to 45% solvent B (80% acetonitrile, 0.1% (v/v) formic acid) and directed through a 3 nL UV detector flow cell. LC was interfaced directly with a 3-D high capacity ion trap mass spectrometer (Esquire HCTplus™, Bruker Daltonics, Bremen, Germany) via a low-volume (50 μl/min maximum) stainless steel nebuliser (Agilent Technologies, Wokingham, UK; cat. no. G1946-20260) and ESI. Parameters for tandem MS analysis were set as previously described [40]. Technical controls included BSA standard and a blank gel slice.

2.4. Database mining

Deconvoluted MS/MS data were searched against an annotated cognate chromosomal and plasmid C. jejuni 81-176 protein database derived from genomic sequences available at the National Centre for Biotechnology Information (Genbank) (Table 1) and the NCBI nr database, using MASCOT software (Matrix Science, London, UK) [42]. Analysis was performed in accordance with published guidelines [43]. To this end, fixed and variable modifications selected were carbamidomethyl (C) and oxidation (M) respectively and mass tolerance values for MS and MS/MS were set at 1.5 Da and 0.5 Da respectively, permitting one missed cleavage. Positive protein identifications were based on recognition of a minimum of two peptides, each with an unbroken series of four or more “b” or “y” ions [40]. Deconvoluted MS/MS data in .mgf (Mascot Generic Format) were imported into ProteinScape™ proteomics data analysis software (Bruker Daltonics, Bremen, Germany) which compiles data from all gel slices utilising the MASCOT search algorithm (Matrix Science, London, UK). The protein content of individual gel slices was established using the “protein search” feature of ProteinScape™, whilst separate compilations of the proteins contained in all 25 gel slices of each of the three biological replicates were produced using the “protein extractor” feature of the software. Data was searched specifying Trypsin and Trypsin/P. Spectra used for protein identifications were re-searched against the entire NCBI nr database to ensure accurate peptide assignments.

2.5. Amino acid sequence comparison tools

Protein sequences were compared by two methods. First, an in-house Java application which takes a single reference genome in protein multi-Fasta and annotation information in .pft formats (http://www.ncbi.nlm.nih.gov/sites/Genome), plus several user-defined comparison genomes in protein multi-Fasta format. This tool uses BLASTP [44] to define putative orthologs between reference and comparison genomes and displays a graphical alignment of these proteins showing alignment score, metadata and an amino acid alignment. The Campylobacter genomes examined are listed in Table 1. NCBI genome datasets comprised (i) Epsilonproteobacteria class, excluding the Campylobacter genus, and (ii) all bacterial genomes excluding the Epsilonproteobacteria class. Second, Smith–Waterman global alignments [45] were carried out between protein pairs. Smith–Waterman scores were normalised using the score for identical sequences as reference (100%) to take account of the disparity in protein length. Scores were displayed as a heat map.

2.6. Bioinformatic tools

The LipoP 1.0 server (http://services.cbu.uib.no/tools/bomp) and Phyre2 was used to predict N-terminal signal peptides in lipoproteins [46]. The β-barrel outer membrane protein predictor (BOMP) server (http://services.cbu.uib.no/tools/bomp) and Phyre2

---

**Table 1**

<table>
<thead>
<tr>
<th>Genus</th>
<th>Classification</th>
<th>NCBI Genome Data</th>
<th>Genbank Accession</th>
</tr>
</thead>
</table>
Table 1 – Genomes used in this study.

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>Source</th>
<th>Accession</th>
<th>Ref.</th>
<th>Genome status</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. jejuni subsp. jejuni 81-176</td>
<td>Contaminated milk</td>
<td>CP000538</td>
<td>NC_008787</td>
<td>Complete</td>
</tr>
<tr>
<td>C. jejuni subsp. jejuni 81-176 pVir</td>
<td></td>
<td>CP000550</td>
<td>NC_008770</td>
<td>Complete</td>
</tr>
<tr>
<td>C. jejuni subsp. jejuni 81-176 pTet</td>
<td></td>
<td>CP000549</td>
<td>NC_008790</td>
<td>Complete</td>
</tr>
</tbody>
</table>

Subject sequences
- C. jejuni subsp. jejuni 1336
- C. jejuni subsp. jejuni 260.94
- C. jejuni subsp. jejuni 327
- C. jejuni subsp. jejuni 414
- C. jejuni subsp. jejuni 81116
- C. jejuni subsp. jejuni 84-25
- C. jejuni subsp. jejuni CF93-6
- C. jejuni subsp. jejuni CG8421
- C. jejuni subsp. jejuni CG8486
- C. jejuni subsp. jejuni DFVF1099
- C. jejuni subsp. jejuni HB93-13
- C. jejuni subsp. jejuni IA3902
- C. jejuni subsp. jejuni ICDCJJ07001
- C. jejuni subsp. jejuni M1
- C. jejuni subsp. jejuni NCTC11168
- C. jejuni subsp. jejuni RM1221
- C. jejuni subsp. jejuni S3
- C. jejuni subsp. doylei 269.97
- C. coli RM2228
- C. coli JV20
- C. upsaliensis RM3195
- C. iari RM2100
- C. curvus 525.92
- C. concisus 13826
- C. fetus 82-40
- C. hominis ATCC BAA-381

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>Source</th>
<th>Accession</th>
<th>Ref.</th>
<th>Genome status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Protein Homology/analogY Recognition Engine V 2.0) (http://www.sbg.bio.ic.ac.uk/phyre2/), were used to predict the presence of β-barrels in peptide sequences [47,48]. Functional domains and motifs were identified in protein sequence using Interproscan server (www.ebi.ac.uk/interpro/) using all available methods [49].

3. Results and discussion

3.1. LC–ESI-MS/MS identification of C. jejuni strain 81-176 Sarkosyl-insoluble proteins

Strict confidence criteria were used when assigning peptide and protein identities [43]. To ensure reproducibility, only proteins represented by two peptides in at least two of three replicates are presented here (Table 2). Sequence coverage was determined using only top ranking peptides. Using the Trypsin/P setting of ProteinScape™, which allows detection of peptides generated by occasional trypsin digestion between lysine or arginine and proline [50], 8 additional peptides were identified, including one which defined l-lactate permease. All matches are found in Tables S1–S6 of the supplementary data.

Applying those criteria, 47 proteins were identified confidently within the Sarkosyl-insoluble material. These comprised several previously characterised OM proteins including the adhesins CadF and FlpA, the MapA antigen and a number of proteins with known roles in small molecule transport including the porins PorA and CJ81176 (Omp50), ABC transporter components PE81, CjaA and CjaC and the multi drug efflux pump component CmeC. Two iron acquisition proteins ChuA and CfrB were also identified as well as the membrane associated flagellar component FlgI, the flagellar hook protein (FlgE), hook-associated protein (FlgH) and both flagellin proteins, FlaA and FlaB, which are enriched under the same conditions as OM proteins [22].
### Table 2 – Proteins identified by LC–ESI-MS/MS analysis of *C. jejuni* 81-176 OM. Number of non-redundant peptides and sequence coverage are given for each biological replicate.

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>NCBI annotation</th>
<th>Gene</th>
<th>No. of peptides</th>
<th>% Coverage</th>
<th>Mr</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJ81176.0019</td>
<td>Conserved hypothetical protein</td>
<td>–</td>
<td>6</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0025</td>
<td>Flagellar hook protein FlgE</td>
<td>FlgE</td>
<td>35</td>
<td>65.9</td>
<td>31</td>
</tr>
<tr>
<td>CJ81176.0056</td>
<td>L-Asparaginase</td>
<td>AnsA</td>
<td>9</td>
<td>41.4</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0067</td>
<td>Gamma-glutamyltransferase</td>
<td>Ggt</td>
<td>8</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0113</td>
<td>L-Lactate peremase</td>
<td>LctP</td>
<td>3</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0116</td>
<td>Cytolethal distending toxin, subunit A</td>
<td>cdtA</td>
<td>3</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0124</td>
<td>Lipoprotein, putative</td>
<td>–</td>
<td>4</td>
<td>12.8</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0125</td>
<td>Lipoprotein, putative</td>
<td>–</td>
<td>3</td>
<td>28.7</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0126</td>
<td>Lipoprotein, putative</td>
<td>–</td>
<td>10</td>
<td>53.1</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0127</td>
<td>Hypothetical protein</td>
<td>–</td>
<td>8</td>
<td>29.7</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0148</td>
<td>Peptidoglycan-associated lipoprotein Omp18</td>
<td>–</td>
<td>5</td>
<td>33.9</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0164</td>
<td>Outer membrane protein, OMP85 family</td>
<td>sodB</td>
<td>3</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0199</td>
<td>Translation elongation factor 7u</td>
<td>Tuf</td>
<td>4</td>
<td>19.5</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0205</td>
<td>Superoxide dismutase, Fe</td>
<td>SodB</td>
<td>3</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0356</td>
<td>Antioxidant, AhpC/Tsa family</td>
<td>AhpC</td>
<td>4</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0388</td>
<td>RND efflux system, outer membrane lipoprotein CmeC</td>
<td>CmeC</td>
<td>7</td>
<td>22.6</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0419</td>
<td>Lipoprotein, putative</td>
<td>–</td>
<td>4</td>
<td>18.7</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0430</td>
<td>Lipoprotein, putative</td>
<td>–</td>
<td>3</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0471</td>
<td>TonB-dependent receptor, putative, degenerate</td>
<td>CfrB</td>
<td>5</td>
<td>10.1</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0499</td>
<td>Translation elongation factor 7u</td>
<td>Tuf</td>
<td>4</td>
<td>19.5</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0586</td>
<td>Conserved hypothetical protein</td>
<td>–</td>
<td>5</td>
<td>20.1</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0641</td>
<td>Nonheme iron-containing ferritin</td>
<td>Ftn</td>
<td>9</td>
<td>62.3</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0710</td>
<td>Flagellar L-ring protein FlgH</td>
<td>FlgH</td>
<td>2</td>
<td>15.1</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0757</td>
<td>CjaC protein</td>
<td>CjaC</td>
<td>8</td>
<td>47.4</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0800</td>
<td>Thiol peroxidase</td>
<td>Tpx</td>
<td>3</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0894</td>
<td>Flagellin family protein</td>
<td>FlgL</td>
<td>5</td>
<td>10.7</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0928</td>
<td>Amino acid ABC transporter, periplasmic amino acid-binding protein PEB1</td>
<td>PEBA</td>
<td>5</td>
<td>30.5</td>
<td></td>
</tr>
<tr>
<td>CJ81176.0974</td>
<td>Conserved hypothetical protein</td>
<td>–</td>
<td>6</td>
<td>53.5</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1001</td>
<td>CjaA protein</td>
<td>CjaA</td>
<td>8</td>
<td>36.2</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1016</td>
<td>Conserved hypothetical protein</td>
<td>–</td>
<td>2</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1045</td>
<td>Lipoprotein, putative</td>
<td>–</td>
<td>5</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1048</td>
<td>Outer membrane lipoprotein MapA</td>
<td>MapA</td>
<td>7</td>
<td>40.2</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1108</td>
<td>Lipoprotein, putative</td>
<td>–</td>
<td>2</td>
<td>26.3</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1124</td>
<td>Methyl-accepting chemotaxis protein</td>
<td>PmbA</td>
<td>5</td>
<td>30.9</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1126</td>
<td>Organic solvent tolerance protein, putative</td>
<td>PmbA</td>
<td>5</td>
<td>30.9</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1127</td>
<td>Major outer membrane protein</td>
<td>PorA</td>
<td>19</td>
<td>66.3</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1195</td>
<td>Fibronectin type III domain protein</td>
<td>FBPB</td>
<td>14</td>
<td>47.3</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1339</td>
<td>Flagellin</td>
<td>Flagellin</td>
<td>29</td>
<td>66.5</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1337</td>
<td>Serine protease, subtilase family</td>
<td>Serine</td>
<td>3</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1374</td>
<td>Lipoprotein, VacJ family</td>
<td>VacJ</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1471</td>
<td>Fibronectin-binding protein</td>
<td>CadF</td>
<td>11</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1519</td>
<td>Bacterioferritin, putative</td>
<td>–</td>
<td>5</td>
<td>47.7</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1601</td>
<td>TonB-dependent heme receptor</td>
<td>Chua</td>
<td>7</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>CJ81176.1690</td>
<td>Ribosomal protein S8</td>
<td>RpsH</td>
<td>3</td>
<td>35.9</td>
<td></td>
</tr>
<tr>
<td>CJ81176.pVir0002</td>
<td>VirB9</td>
<td>–</td>
<td>1</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>CJ81176.pVir0048</td>
<td>Conserved hypothetical protein</td>
<td>–</td>
<td>4</td>
<td>39.2</td>
<td></td>
</tr>
</tbody>
</table>

* Sequences with re-assigned start codons were used.

Three lipoproteins, which are encoded by a locus of three genes, were also identified (CJJ81176.0124, CJ81176.0125 and CJ81176.0126). Oakland et al. [51] recently characterised homologues of these lipoproteins in strain NCTC11168 (Cj0089, Cj0090, and Cj0091). Few autotransporter proteins are annotated in *C. jejuni* yet they are common features of Gram-negative bacterial outer membranes. In this study we detected the expression of one putative autotransporter, the serine protease CJ81176.1367. Bioinformatic analysis of this protein shows several features of autotransporters.
including the β-strand repeat (IPR013425) and β-barrel domain (IPR005546).

The roles of Omp18 (CJJ81176_0148), Omp85 (CJJ81176_0164) and the putative organic solvent tolerance protein (CJJ81176_1268) have not been investigated in C. jejuni although homologues of these proteins in other bacteria are involved in maintenance of membrane integrity and biogenesis. Omp18 is also known as peptidoglycan associated protein (Pal), which is a component of the well-characterised Tol-Pal envelope complex [52]. Within Campylobacter genomes, Omp85 is the closest homologue of BamA, which in other Gram-negative bacteria is the OM component of the β-barrel assembly complex [25]. The protein annotated as a putative organic solvent tolerance protein is the closest Campylobacter homologue to E. coli LptD (LPS transport) [53]. Disruption of a Helicobacter pylori homologue, OstA (29% sequence identity with CJJ81176_1268) resulted in an increase in membrane permeability, susceptibility to hydrophobic and β-lactam antibiotics and sensitivity to organic solvents [54] and possibly performs a similar role in C. jejuni.

Eleven further gene translation products were also identified, including nine uncharacterised proteins of chromosomal origin and two plasmid encoded proteins – VirB9, and the hypothetical protein CJJ81176_pVir0048.

Several proteins identified in this study are by consensus considered to be cytoplasmic, although homologues of superoxide dismutase (SOD), thiol peroxidase (Tpx), bacterioferritin (Dps), gamma-glutamyltransferase (GGT) and L-asparaginase have been localised to the OM in the closely related H. pylori [55–58]. Esposito et al. [59] speculated that the localisation of SOD in H. pylori may be due to an extended C-terminal tail, a feature shared with Campylobacter SOD. Protein CJJ81176_1519 is annotated as a putative bacterioferritin although was characterised by Ishikawa et al. [60] as a Dps (DNA protection during starvation) protein, a protein class which is widely distributed in bacterial species and are members of the ferritin superfamily. The protein with highest sequence similarity to CJJ81176_1519 outside Campylobacter spp. is NapA of H. pylori, a surface exposed adhesin [61]. GGT, a component of the antioxidant glutathione pathway, is required for persistent colonisation of the avian intestine by C. jejuni 81116 [62] although GGT is not present in all C. jejuni isolates, including strains NCTC11168 and RM1221. Elongation factor Tu (EF-Tu) is central to protein synthesis in the cytoplasm although it is now considered to play a variety of roles within the cell and has been localised to the OM of Gram-negative bacteria, including Neisseria meningitidis [63] as well as the surface of Mycoplasma pneumoniae and Lactobacillus johnsonii where it is thought to contribute to host adhesion [64–66]. Thus, as increasingly shown for other bacteria, C. jejuni may express non-classically associated OM “moonlighting” proteins on its surface with potential roles in infection.

### 3.2. Suggested re-annotation of C. jejuni strain 81-176 open reading frames

The annotation of CJJ81176_1108 as a putative lipoprotein and the presence of β-barrel domain of CJJ81176_1268 suggest OM association of both proteins although their annotated amino acid sequences lack signal peptides. Closer examination of the genome sequences upstream from the annotated sequences within the 81-176 genome (NC_008787) for CJJ81176_1108 and CJJ81176_1268 revealed alternative translational start sites for these sequences, resulting in an additional N-terminal 56 and 7 amino acids respectively. These alternative N-terminal sequences contain intact lipoprotein signal peptides and are annotated as translational start sites for orthologues in other C. jejuni subsp. jejuni genomes in NCBI databases. Furthermore, MS data searched against the revised sequence for CJJ81176_1108 identified a peptide indicating expression of this region of the protein. This evidence suggests that the translational start codons for accession entries gi|121612654 and gi|121504146 are currently mis-annotated; the suggested revised N-terminal sequences are shown Figure S1 of the supplementary data.

### 3.3. Distribution of proteins across Campylobacter species

We used in-house software to examine conservation and sequence similarity of a selected panel of characterised and uncharacterised OM proteins within Campylobacter genomes. A graphical representation of sequence similarity is provided, simultaneously facilitating the identification of conserved determinants that may be important in Campylobacter fitness and pathogenicity, heterogenous coding sequences, and highlights the presence of potential species/strain-specific markers. Fig. 1 shows comparisons for selected OM proteins against Campylobacter species and strains and selected NCBI datasets. This analysis demonstrates that the majority of OM proteins are conserved amongst C. jejuni subsp. jejuni strains, with the exception of the serine protease CJJ81176_1367 and hypothetical protein CJJ81176_0019 (discussed in more detail in the next section). The results also indicate that there is little sequence similarity between C. jejuni OM proteins and proteins within Campylobacter spp. other than C. coli. The majority of OM proteins shared high sequence similarity with proteins present within both strains of C. coli although the lipoprotein CJJ81176_0125, CJJ81176_1185 (Omp50), the hypothetical protein CJJ81176_0127 and the serine protease CJJ81176_1367 are absent. Several OM proteins of strain 81-176 were revealed to have high sequence similarity (up to 91%) with C. upsaliensis and/or C. lari proteins, although many OM proteins were divergent or absent. Only a few protein sequences are conserved in Campylobacter spp. curvus, concisus, fetus and hominis and the remaining proteins are either divergent or absent completely. Results of BLASTP searches of selected NCBI databases reveals presence of conserved proteins within Epsilonproteobacteria class (excluding the Campylobacter genus), as well as more widely amongst bacteria.

### 3.4. Analysis of sequence variability of OM proteins within C. jejuni isolates

Given the reported genetic variation within Campylobacter strains and species [24], conservation and sequence similarity of OM proteins in C. jejuni genomes were further investigated. The Smith–Waterman algorithm was used to generate alignments, instead of the time-optimised BLAST algorithm, and normalised scores were displayed as a heat map to allow...
Fig. 1 – Comparison of C. jejuni 81-176 OM protein amino acid sequences identified by LC–ESI-MS/MS with amino acid sequences derived from 27 selected Campylobacter genomes and specified NCBI genomic datasets, using the BLASTP algorithm. Reciprocal best hits are identified and represented graphically with the strength of homology shown as shaded rectangles. Colour coding of C. jejuni 81-176 query sequences corresponds to Cluster of Orthologous Groups (COGs), allocated by National Center for Biotechnology Information (NCBI) which indicates predicted function.

greater discrimination (Fig. 2). Amino acid sequences for CadF, Cj81176_1268, Omp18/Pal, Cj81176_0126, and the uncharacterised proteins Cj81176_0430, Cj81176_0974, Cj81176_1016, Cj81176_1108, and Cj81176_0419 are all highly conserved across C. jejuni subsp. jejuni isolates, which suggests that these are essential membrane proteins. Omp50, PEB1, and FlpA were also found to be conserved, although some sequence variations exist, all of which are predicted to be a result of assignment of alternative start codons. Alternatively assigned start codons are shown in Table S7 of the supplementary data.

Several proteins showed divergence across strains with respect to sequence heterogeneity as well as presence/absence. PorA, ChuA, and Cj81176_1185 (Omp50) homologues exist in all C. jejuni subsp. jejuni strains but the Smith–Waterman scores indicate divergence in sequence; closer inspection of sequence alignments shows all three proteins contain regions of variability. This is of particular interest for Omp50, which was recently shown to coordinate capsule formation via a tyrosine phosphorylation cascade [17]. Whether this sequence variability influences protein function warrants further study. The variability of PorA is well documented and has been exploited in the development of molecular epidemiology typing methods [67–70]. The serine protease autotransporter Cj81176_1367 is absent in strains M1, 81116, and 327. Additionally, a divergent sequence is present within strains ICDCJC07001 and 260.94, although this sequence shares stronger sequence similarity with Cj81176_1367, an additional sequence within the C. jejuni 81-176 genome. Cj81176_0019 is divergent in strains NCTC11168, DFVF1099, and cotx52_40. C. hominis ATCC BAA-381 c-proteobacteria

non-c-proteobacteria

Cluster of Orthologous Groups

- Posttranslational modification, protein turnover, chaperones
- Gene regulation, transcription
- Inorganic ion transport and metabolism
- Carbohydrate transport and metabolism
- Signal transduction mechanisms
- General function prediction only
- Function unknown
- Not available (forward)
- Not available (reverse)

epidemiology typing methods [67–70]. The serine protease autotransporter Cj81176_1367 is absent in strains M1, 81116, and 327. Additionally, a divergent sequence is present within strains ICDCJC07001 and 260.94, although this sequence shares stronger sequence similarity with Cj81176_1367, an additional sequence within the C. jejuni 81-176 genome. Cj81176_0019 is divergent in strains NCTC11168, DFVF1099, and cotx52_40. C. hominis ATCC BAA-381 c-proteobacteria

non-c-proteobacteria

Cluster of Orthologous Groups

- Posttranslational modification, protein turnover, chaperones
- Gene regulation, transcription
- Inorganic ion transport and metabolism
- Carbohydrate transport and metabolism
- Signal transduction mechanisms
- General function prediction only
- Function unknown
- Not available (forward)
- Not available (reverse)

Several proteins showed divergence across strains with respect to sequence heterogeneity as well as presence/absence. PorA, ChuA, and Cj81176_1185 (Omp50) homologues exist in all C. jejuni subsp. jejuni strains but the Smith–Waterman scores indicate divergence in sequence; closer inspection of sequence alignments shows all three proteins contain regions of variability. This is of particular interest for Omp50, which was recently shown to coordinate capsule formation via a tyrosine phosphorylation cascade [17]. Whether this sequence variability influences protein function warrants further study. The variability of PorA is well documented and has been exploited in the development of molecular
result in altered virulence, metabolic capability or survivability. Some truncations are of considerable size, and in some cases the resulting protein sequences are annotated as pseudogenes although annotation of some ORFs in different strains is inconsistent. Some proteins are found to be truncated in many strains; truncations of CjaC are found within five strains, all of which appear to be due to frameshifts. Diversity within CjaC has been reported previously [71]. Many truncations were observed to occur as a result of insertions/deletions within homopolymeric tracts. Notably, the tracts are A or T rather than G or C which is more commonly documented in Campylobacter spp. Although it is tempting to speculate that these A or T homopolymeric tracts may allow a rapidly adaptive and reversible alteration in OM protein composition and a potential survival/adaptation strategy, it is more likely that many of these indels are a result of the high error rate in high throughput sequencing of homopolymeric regions.

Iron availability is a critical requirement for bacteria, resulting in diverse uptake mechanisms. Several iron acquisition systems have been identified in C. jejuni, many with OM components. Variation in distribution of these systems is well documented and is discussed in a review by Miller et al. [18]. Sequence analysis of the ferric enterobactin receptor CfrB also reveals a degree of variation within coding sequences for this protein. An intact coding sequence for CfrB is present within strains ICDCJ07001, 260.94, HB93-13, 81116, M1, and 327. Frameshifts resulting in truncated proteins in strains 84-25, DFVF1099 and CF93-6 were also observed. Manual inspection revealed that within NCBI protein sequence databases for strains NCTC11168, RM1221, S3, IA3902, CG8421 and CG8486, the protein with closest sequence similarity to CfrB is the iron binding protein CfrA, which is not found within the strain 81-176 genome. Despite the absence of annotated protein sequences within the databases held at NCBI for several strains, homologous nucleotide sequences similar to cfrB are present in all genomes, although several sequences contain indels which results in their exclusion from annotated protein sequence databases. A summary of frameshifts and resulting outcomes is shown in Table S9 of the supplementary data. Significantly, all genomes with truncated CfrB proteins contain an intact coding sequence for CfrA and conversely, genomes with highly similar CfrB sequences lack CfrA indicating a greater conservation in strains without the alternative CfrA system.

Divergence in environmental isolates and C. jejuni subsp. doylei 269.97 is also apparent. Strain 414 and, to a lesser extent, strain 1336 show divergence in several proteins which are conserved in other C. jejuni subsp. jejuni strains. Fig. 2 shows that several proteins conserved within C. jejuni subsp. jejuni isolates are absent from C. jejuni subsp. doylei. C. jejuni subsp. doylei is associated with both gastritis and enteritis and is more commonly bacteraemia, particularly in paediatric patients, and has been frequently isolated from blood cultures [72].

4. Concluding remarks

In 2008, Cordwell et al. carried out a comprehensive investigation of the membrane compartment of C. jejuni although this analysis was not restricted to the OM as it included the inner membrane and periplasmic space [30]. Membrane-associated proteins are typically hydrophobic, thus in order to investigate the proteome of the OM compartment of C. jejuni 81-176, we enriched Sarkosyl-insoluble bacterial proteins and applied a gel-based proteomics approach. Characterisation of the OM
compartment is paramount to understanding pathogenicity, and it is also important to establish the distribution of these factors within strains. Equally, proteins conserved within disease-causing species which are absent in non-pathogenic species suggest a potential requisite for pathogenicity. We also investigated conservation of a panel of OM proteins within selected genomic datasets. Whilst it is apparent that individual isolates may contain a diversity of proteins within the OM, it is also clear that a number of proteins are conserved; a large proportion of these remain functionally uncharacterised, however.

The similarities between C. jejuni and H. pylori OM emphasise a requirement for further characterisation of Epsilonproteobacteria OM, for which common biological processes are already known to differ from those of the well-studied Enterobacteriaceae [73]. Since protein content varies both quantitatively and qualitatively in response to environmental cues, future efforts will examine homogeneity and heterogeneity of protein expression within and between strains under different culture conditions and phases of growth, in response to specific stimuli and upon interaction with hosts. The results presented here reinforce the notion that C. jejuni should be considered as heterogeneous bacteria with potential for considerable variability at the genotypic, phenotypic and ultimately pathogenic levels.

Conflict of interest

All the authors have reported that they have received grants from Scottish Government/RESAS and BBSRC during the conduct of the study.

Transparency document

The Transparency document associated with this article can be found in the online version.

Acknowledgements

Thanks to Kevin Mclean for useful discussions regarding interpretation of mass spectrometry data and Thorsten Forster for assistance with graphics. The work was supported by Scottish Government Rural and Environment Science and Analytical Services Division Programme funding and BBSRC funding (Reference No. BB/F020988/1).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.euprot.2014.06.003.

REFERENCES


subcellular fractionation and nano-LC FT-ICR MS analysis. J Proteome Res 2006;5:3197–204.


