Fingerprinting of *Campylobacter* Species and Strains by Mass Spectrometric Identification of Protein Biomarkers

Clifton K. Fagerquist*[1], Leslie A. Harden[1], William G. Miller[1], Derrick E. Fouts[2], Karen E. Nelson[2], Robert E. Mandrell[1]

[1]Western Regional Research Center, Agricultural Research Service, USDA, 800 Buchanan Street, Albany, CA. [2]The Institute for Genomic Research, Rockville, MD.

Overview

Purpose

- Validate the use of MALDI-TOF-MS for biomarker analysis of bacterial cell lysates by unambiguous identification of select protein biomarkers;
- Determine whether a single protein biomarker can discriminate *Campylobacter* species and strains;

Method

- MALDI-TOF-MS analysis of *Campylobacter* cell lysates;
- HPLC of cell lysates to isolate proteins;
- Identify HPLC fractions containing protein biomarkers using MALDI-TOF-MS;
- 1-D PAGE of HPLC fractions containing biomarkers;
- In-gel digestion of gel bands;
- LC/MS/MS of digests and database searching for unambiguous identification of select protein biomarkers;
- Confirm experimental biomarker MW against theoretical database protein MW.

Results

- Unambiguous identification of a prominent protein biomarker is HUP: a DNA-binding protein;
- The HUP protein biomarker can be used to discriminate species and strains of *Campylobacter*

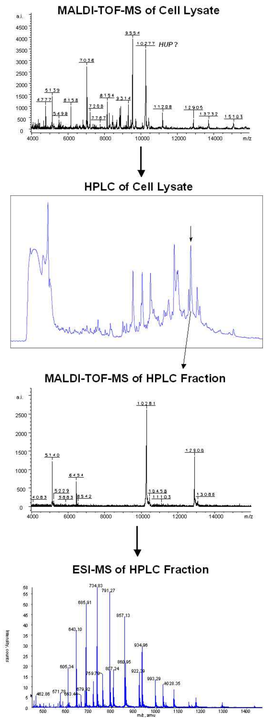
Introduction

- **Bacterial illness attributed to *Campylobacter* accounts for a high percentage of the estimated 77 million yearly incidents of food-borne illness in the USA.**
- **Seven *Campylobacter* identified to date have been identified as potential human pathogens: *C. jejuni, C. coli, C. lari, C. fetus, C. upsaliensis, C. sputorum* and *C. concisus*.**
- **Although the majority of cases of campylobacteriosis involve short-term symptoms typical of bacterial illness, approximately 1 in a 1000 cases may be linked to a subsequent occurrence of Guillain-Barré Syndrome, the most commonly occurring paralytic disease in the USA.**
- **There is a need for a simple, rapid, rugged, sensitive and specific technique for identifying *Campylobacter* species and strains.**
- **MALDI-TOF-MS analyses of whole cell lysates provide a relatively simple method of species and strain discrimination by identification of unique protein biomarkers.**

Experimental

- *Campylobacter* cells were harvested after several days of growth on plates;
- Cell dispersion/lysis was accomplished using a simple solvent extraction solution (67% water, 33% acetonitrile, 0.1% TFA) followed by bead-beating (0.1 mm zirconia/silica beads);
- MALDI-TOF-MS analysis involved suspension of the protein extract in a ferulic acid matrix and analysis of ions in the 4kD-16kD mass range in reflectron mode using a Bruker Reflex II mass spectrometer;
- Prominent biomarker ions useful for discrimination of *Campylobacter* species and strains were isolated by RP-HPLC and 1-D SDS PAGE;
- Excised gel bands were subjected to in-gel digestion with trypsin (or chymotrypsin) and peptides were analyzed by nano-LC/MS/MS analysis on a Q-STR Pulsar I with nano-ESI;
- Genomic sequencing of *Campylobacter* species provided theoretical protein sequences and molecular weights;
- Database searching of the MS/MS spectra provided unambiguous identification of the protein biomarker;
- Experimental MW of extracted protein biomarkers (from deconvolution of ESI-MS) compared against theoretical protein MW.

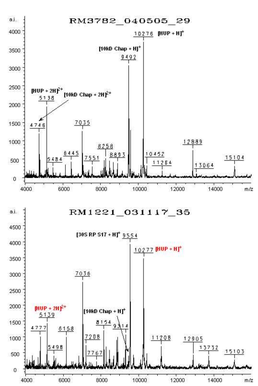
Experimental Schematic



∴ **Protein biomarker unambiguously identified as HUP.**

C. jejuni

(Spectra arranged by increasing *HUP* MW, **Red** indicates biomarker unambiguously ID by MS/MS.)



Sequence coverage from LC/MS/MS of in-gel digestions (trypsin) of gel band from cell lysate of RM1221 (*C. jejuni*) using Global Proteome Machine software.

log(e) = -147.1 RM_1221_hup_[DNA-binding_protein_HU_homolog_]

1 MTKADFISLVAQTAGLTKKDATTATDAVISTITDVLAKGDSISFIGFGTFSTQERAAREA 60

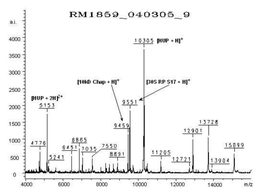
61 RVPSTGKTIKVPATRVAKFKVGKNLKEAVAKASGKKKK 98

Sequence coverage from LC/MS/MS of in-gel digestion (chymotrypsin) of gel band from cell lysate of RM1221 (*C. jejuni*) using Global Proteome Machine software.

log(e) = -240.5 RM_1221_hup_[DNA-binding_protein_HU_homolog_]

1 MTKADFISLVAQTAGLTKKDATTATDAVISTITDVLAKGDSISFIGFGTFSTQERAAREA 60

61 RVPSTGKTIKVPATRVAKFKVGKNLKEAVAKASGKKKK 98



Sequence coverage from LC/MS/MS of digestion (chymotrypsin) of gel slice X using Global Proteome Machine software.

log(e) = -240.1 RM_1221_hup_[DNA-binding_protein_HU_homolog_]

1 MTKADFISLVAQTAGLTKKDATTATDAVISTITDVLAKGDSISFIGFGTFSTQERAAREA 60

61 RVPSTGKTIKVPATRVAKFKVGKNLKEAVAKASGKKKK 98

Sequence coverage from LC/MS/MS of digestion (trypsin) of gel slice X using Global Proteome Machine software.

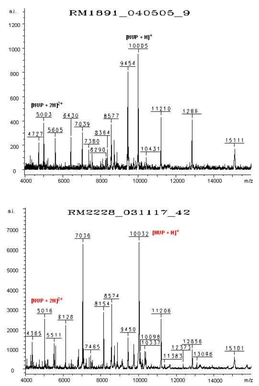
log(e) = -137.3 RM_1221_hup_[DNA-binding_protein_HU_homolog_]

1 MTKADFISLVAQTAGLTKKDATTATDAVISTITDVLAKGDSISFIGFGTFSTQERAAREA 60

61 RVPSTGKTIKVPATRVAKFKVGKNLKEAVAKASGKKKK 98

C. coli

(Spectra arranged by increasing *HUP* MW, **Red** indicates biomarker unambiguously ID by MS/MS.)



Sequence coverage from LC/MS/MS of in-gel digestion (trypsin) of gel band from cell lysate of RM2228 (*C. coli*) using Global Proteome Machine software.

log(e) = -152.2 RM_2228_hup_[DNA-binding_protein_HU_homolog_]

1 MTKADFISQVAQTAGLTKKDATAATDAVISTITEVLAKGDSVSFIGFGTFSTAERAAREA 60

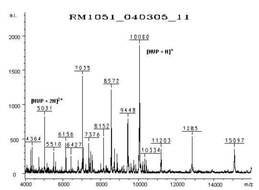
61 RVPSTGKTIKVPATRVAKFKVGKNLKDAVAKAKKKK 96

Sequence coverage from LC/MS/MS of in-gel digestion (chymotrypsin) of gel band from cell lysate of RM2228 (*C. coli*) using Global Proteome Machine software.

log(e) = -264.4 RM_2228_hup_[DNA-binding_protein_HU_homolog_]

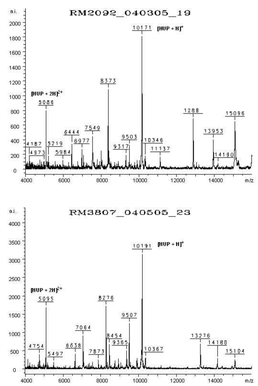
1 MTKADFISQVAQTAGLTKKDATAATDAVISTITEVLAKGDSVSFIGFGTFSTAERAAREA 60

61 RVPSTGKTIKVPATRVAKFKVGKNLKDAVAKAKKKK 96



C. upsaliensis

(Spectra arranged by increasing *HUP* MW, **Red** indicates biomarker unambiguously ID by MS/MS.)



Sequence coverage from LC/MS/MS of in-gel digestion (trypsin) of gel band from cell lysate of RM3195 (*C. upsaliensis*) using Global Proteome Machine software.

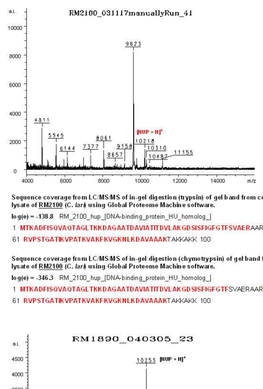
log(e) = -102.8 RM_3195_hup_[DNA-binding_protein_HU_homolog_]

1 MTKADFFSQVAQRAELTKKDAGAATDAVISTITEVLAKGDSRSFKGFGTFSTTERAAREA 60

61 RVPSTGKTIKVPATRVAKFKVGKNLKEAVAKAGKKKKK 98

C. lari

(Spectra arranged by increasing *HUP* MW, **Red** indicates biomarker unambiguously ID by MS/MS.)



Sequence coverage from LC/MS/MS of in-gel digestion (trypsin) of gel band from cell lysate of RM2100 (*C. lari*) using Global Proteome Machine software.

log(e) = -130.8 RM_2100_hup_[DNA-binding_protein_HU_homolog_]

1 MTKADFFSQVAQTAGLTKKDAGAATDAVIATITDVLAKGDSRSFKGFGTFSVAERAAREA 60

61 RVPSTGATIKVPATKVAKFKVGKNLKDAVAAAKTAKKAKK 100

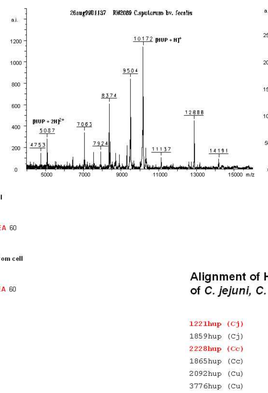
Sequence coverage from LC/MS/MS of in-gel digestion (chymotrypsin) of gel band from cell lysate of RM2100 (*C. lari*) using Global Proteome Machine software.

log(e) = -346.3 RM_2100_hup_[DNA-binding_protein_HU_homolog_]

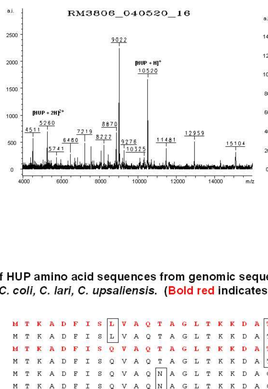
1 MTKADFFSQVAQTAGLTKKDAGAATDAVIATITDVLAKGDSRSFKGFGTFSVAERAAREA 60

61 RVPSTGATIKVPATKVAKFKVGKNLKDAVAAAKTAKKAKK 100

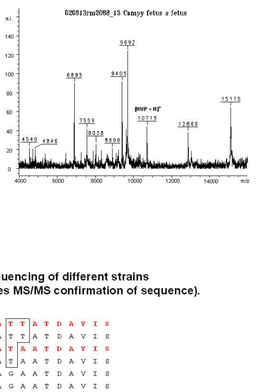
C. sputorum



C. concisus



C. fetus



Alignment of HUP amino acid sequences from genomic sequencing of different strains of *C. jejuni, C. coli, C. lari, C. upsaliensis*. (Bold red indicates MS/MS confirmation of sequence).

```
1221hup (Cj)   M T K A D F I S L V A Q T A G L T K K D A T T A T D A V I S
1859hup (Cj)   M T K A D F I S L V A Q T A G L T K K D A T T A T D A V I S
2228hup (Cc)   M T K A D F I S Q V A Q T A G L T K K D A T A A T D A V I S
1865hup (Cc)   M T K A D F I S Q V A Q T A G L T K K D A T A A T D A V I S
2092hup (Cu)   M T K A D F I S Q V A Q N A G L T K K D A G A A T D A V I S
3776hup (Cu)   M T K A D F I S Q V A Q N A G L T K K D A G A A T D A V I S
3195hup (Cu)   M T K A D F I S Q V A Q H A E L T K K D A G A A T D A V I S
4245hup (Cu)   M T K A D F I S Q V A Q N A G L T K K D A G A A T D A V I S
2100hup (Cl)   M T K A D F I S Q V A Q T A G L T K K D A G A A T D A V I A

1221hup (Cj)   T I T D V L A K G D S I S F I G F G T F S T Q E R A A R E A
1859hup (Cj)   T I T D V L A K G D S I S F I G F G T F S T Q E R A A R E A
2228hup (Cc)   T I T D V L A K G D S V S F I G F G T F S T A E R A A R E A
1865hup (Cc)   T I T D V L A K G D S V S F I G F G T F S T T E R A A R E A
2092hup (Cu)   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 A A R E A
3776hup (Cu)   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 A A R E A
3195hup (Cu)   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 A A R E A
4245hup (Cu)   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 A A R E A
2100hup (Cl)   T I T D V L A K G D S I S F I G F G T F S V A E R A A R E A

1221hup (Cj)   R V P S T G K T I K V P A T R V A K F K V G K N L K E A V A
1859hup (Cj)   R V P S T G K T I K V P A T R V A K F K V G K N L K E A V A
2228hup (Cc)   R V P S T G K T I K V P A T R V A K F K V G K H L K D A V A
1865hup (Cc)   R V P S T G K T I K V P A T R V A K F K V G K N L K D A V V
2092hup (Cu)   R V P S T G K T I K V P A T R V A K F K V G K S L K E A V A
3776hup (Cu)   R V P S T G K T I K V P A T R V A K F K V G K N L K E A V A
3195hup (Cu)   R V P S T G K T I K V P A T R V A K F K V G K H L K E A V A
4245hup (Cu)   R V P S T G K T I K V P A T R V V K F K V G K N L K E A V A
2100hup (Cl)   R V P S T G A T I K V P A T K V A K F K V G K H L K D A V A

1221hup (Cj)   - K A S G K K K K
1859hup (Cj)   - K A S K K K K
2228hup (Cc)   - K A - - K K K K
1865hup (Cc)   - K A - - K K K K
2092hup (Cu)   A K A - G K K K K
3776hup (Cu)   A K A - G K K K K
3195hup (Cu)   A K A - G K K K K
4245hup (Cu)   A K A - G K K K K
2100hup (Cl)   A A K T A K K A K K
```

Comparison of the HUP molecular weight from genomic sequencing and the molecular weight of HUP from deconvolution of the ESI charge state envelope of the extracted protein.

Strain	Database MW (in Daltons)	Q-STR MW (in Daltons)	MS/MS Coverage (%)
RM1221 (C.j.)	10274.0	10274.0	81.6 %
RM1859 (C.j.)	10304.0	10304.0	---
RM2228 (C.c.)	10029.7	10029.4	90.6 %
RM1051 (C.c.)	10059.7	10059.5	---
RM3195 (C.u.)	10270.9	10270.9	88.8 %
RM2092 (C.u.)	10171.8	10171.7	---
RM3776 (C.u.)	10198.9	10198.9	---
RM2100 (C.l.)	10210.9	10210.8	90.0%

Conclusions

This work demonstrates that it is possible to…

- **Discriminate species and strains of *Campylobacter* by mass identification of protein biomarkers from cell lysates analyzed by MALDI-TOF-MS;**
- **Isolate prominent protein biomarkers observed in the MALDI-TOF-MS spectra of bacterial lysates and unambiguously identify them by a combination of genomic and proteomic analysis;**
- **Use select protein biomarkers to discriminate between *Campylobacter* species and strains and this discrimination is determined at the bacterial genomic level.**

Future Work

- **Isolate and identify other *Campylobacter* protein biomarkers, e.g. 10 kD chaperonin.**
- **Test other *Campylobacter* species and strains as well as other pathogens.**

Acknowledgements

We wish to thank William H. Vensel, Anna Bates, Guilin Wang, Samarpita Sarkar for their assistance and William F. Haddon for his prior contributions to this work.