Mass Spectral Advances in Diagnostic Microbiology: from Electrophoretic Typing to LC-MS/MS-based Approaches

Min Fang[1], Tom Gaulton[1], Raju Misra[2], Saheer E. Gharbia[2], Haroun N. Shah[1]
Proteomics Research Unit[1], Culture Collections and Genomics Research Unit[2], Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK

INTRODUCTION

In recent years MALDI-TOF-Mass Spectrometry has risen to the forefront of diagnostic microbiology largely because its comprehensive database which enabled its resolution to reach, and often surpass, the current gold standard, 16S rRNA for microbial identification[1]. Furthermore, its simplicity, speed of analysis and low cost has impelled MALDI-MS to gain universal acceptance in clinical laboratories. New developments in data analysis are allowing some species to be subtyped but for other pathogens the limits of resolution have been reached and new approaches are required.

Bacterial-typing techniques, such as SDS-PAGE profiles generated by gel electrophoresis have been used for decades as a basis for electrophoretic typing of closely related microbial species and shows excellent congruence with DNA-based methods[2]. SDS-PAGE in combination with electrospray tandem MS (designated GelC-MS/MS) is frequently used for proteomic analysis[3]. However, its use as a platform for microbial proteotyping has not been systematically investigated. Here, we subjected members of the family *Enterobacteriaceae* including taxonomically indistinguishable species such as *E. coli* and *Shigella* spp. to such analyses to explore the potential of this approach.

The results show that by using the optimised database and proteome profiling MS-generated data, all taxa could be confidently delineated to the species and subspecies levels. Furthermore, in addition to comparative mass spectral analysis, data could be obtained on the biological properties of a strain. Thus, we could confidently identify *E. coli*, characterise strain-specific virulence factors and differentiate Shiga toxin negative from positive strains. This proof of concept study demonstrates that GeLC-MS/MS, which combines the traditional SDS-PAGE platform with LC-MS/MS, has the potential to simultaneously combine strain identification with key pathogenic properties of an isolate and significantly extends the clinical of applications of mass spectrometry in microbiology.

METHODS

A panel of 33 species, comprising 70 enteric bacterial strains (Table 1) were cultured, lysed prior to protein identification by LC-MS/MS using a Thermo LTQ Orbitrap. The MS/MS data were then searched against a series databases using Mascot (Matrix Science, UK) for bottom-up protein identification (Figure 1).

Marker Database Generation
In order to generate marker database, a series of standalone NCBI BlastP search against databases which increased in size: i) genus specific protein sequences; ii) *Enterobacteriaceae* family and iii) NCBInr;

Analysis of an Outbreak *E. coli* 0104 Strains
A selection of pathogenic *E. coli* strains including three isolates from the 2011 German outbreak (H112160280, H112160540 and H112160541), and four strains representative for the two characterised pathovars, EAEC (E99518), EHEC (EDL933 and H10302) and a Non-toxigenic, shigatoxin negative EHEC strain (NCTC12900), were analysed using the bottom proteomics method described in this study.

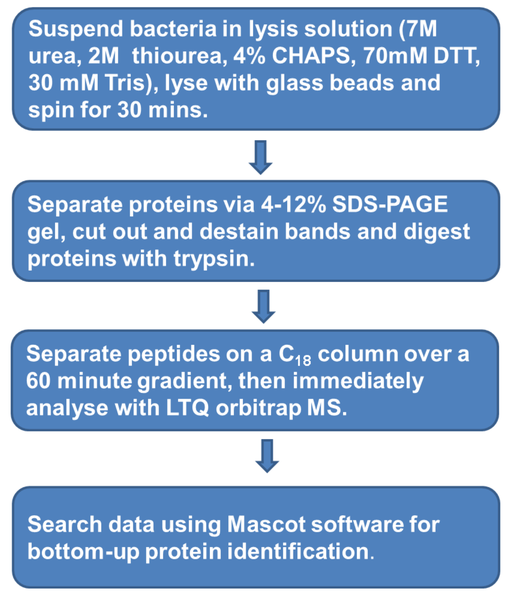

Figure 1. GelC-MS/MS workflow.

RESULTS

- **Database Development**

  A panel of 70 strains from the complex *Enterobacteriaceae* family, including the type strains NCTC09001 and NCTC12985 were analysed using GelC-MS/MS to generate a database. The final database comprised of peptides from a non-redundant list of 20678 protein sequences;

  Analysis of functions and subcellular locations of the parent proteins in the database, demonstrated that the identified proteins varied greatly relating to function and subcellular location (Figure 2).

- ***E. coli* Isolates from Outbreak**

  By using the *Enterobacteriaceae* database generated, all seven test strains were matched and identified as *E. coli*;

  Investigation of the total proteome for each of the seven strains revealed a varied list of proteins, covered a wide range of functional categories, including virulence factors;

  The outbreak strains expressed both EAEC and EHEC features as well as features specific to outbreak strains, including antibiotic resistance, heavy metal resistance and toxin proteins;

  Proteins unique to each pathovar were also identified, indicating that the approach has the potential to delineate pathogenic *E. coli* into subspecies groups;

  Shiga-like toxins were identified from Shiga toxin positive isolates (H112160280, H112160540, H112160541, EDL933 and H10302), which could be used to differentiate Shiga toxin positive strains from Shiga toxin negative strains (E99518 and NCTC12900) (Figure 3).

Enterobacteriaceae

Genus	Species	Number of Strains
Escherichia	*coli*	9
	hermannii	
	vulneris	
	fergusonii	
Salmonella	*enterica*	2
Shigella	*flexneri*	4
	sonnei	
	boydii	
Klebsiella	*pneumoniae*	8
	aerogenes	
Enterobacter	*aerogenes*	7
	cloacae	
	gergoviae	
	sakazakii	
Proteus	*mirabilis*	8
	penneri	
	rettgeri	
	vulgaris	
Providencia	*alcalifaciens*	8
	rettgeri	
	stuartii	
Yersinia	*enterocolitica*	6
	pseudotuberculosis	
Hafnia	*alvei*	3
Morganella	*morganii*	3
Serratia	*marcescens*	3
Campylobacter	*jejuni*	5
	coli	
	fetus	
	hyointestinalis	
	upsaliensis	
Citrobacter	*Freundii*	1
Cronobacter	*sakazakii*	3

Table 1. The 70 strains *Enterobacteriaceae* used in this study.

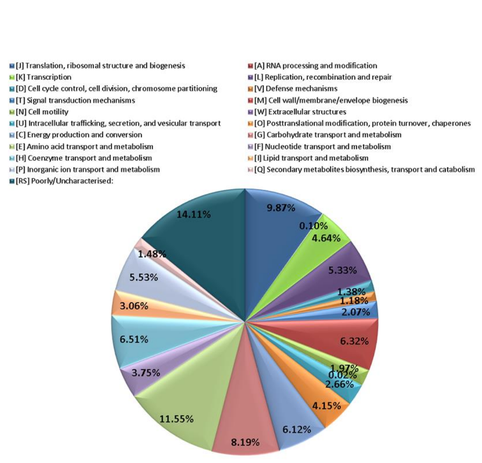

Figure 2. . Distribution of COG functional categories for all parent proteins in the *Enterobacteriaceae* database.

E. coli	Strains	Shigatoxins Identified
Outbreak	H112160280 H112160540 H112160541	Shiga-like toxin II subunit A precursor Shiga-like toxin II subunit B precursor
EAEC (Shigatoxin Negative)	E99518	N/A
EHEC (Shigatoxin Positive)	EDL933 H10302	Shiga-like toxin II subunit A precursor Shiga-like toxin II subunit B precursor
EHEC (Shigatoxin Negative)	NCTC12900	N/A

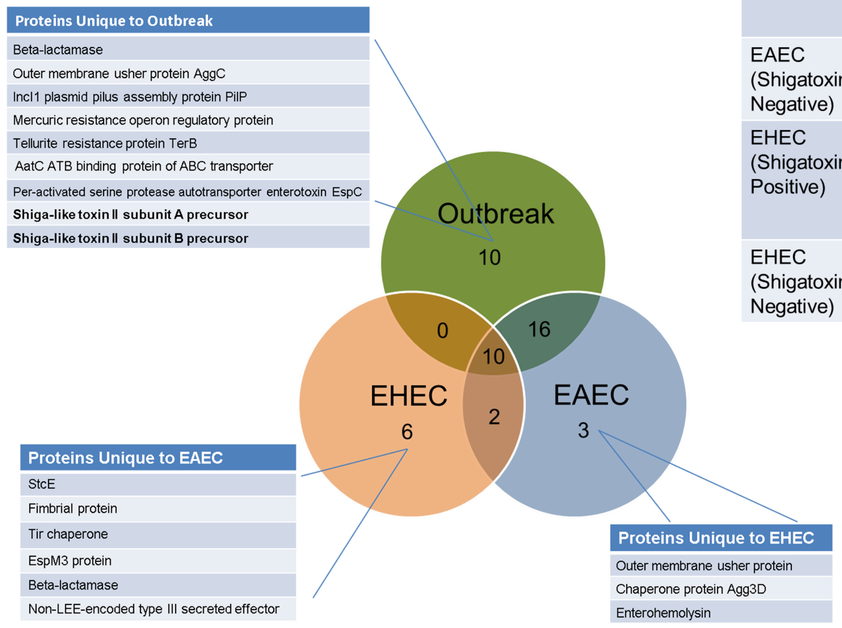

Figure 3. Overlap and concordance between outbreak, EAEC and EHEC isolates.

DISCUSSION

1. This study revealed that genus, species-specific and even strain-specific biomarkers can be deduced for each taxon, including those closely related *Enterobacteriaceae* strains, making it one of the high resolution and high accuracy tools available to date;
2. Bottom-up proteomics is capable of identifying a far greater number of proteins belonging to multi-functional categories, which were used to populate a database that could be used to characterise members of complex *Enterobacteriaceae* family;
3. Protein expression data indicated that the outbreak strains contained not only EAEC and EHEC features, but also unique features to outbreak strains, *i.e.,* antibiotic resistance, heavy metal resistance and toxins, indicating the outbreak strain was a mixed pathovar strain;
4. Identification of Shiga-like toxins enables Shiga-toxin positive strains to be differentiated from Shiga-toxin negative strains.

CONCLUSIONS

1. By using the optimised database and proteome profiling generated-MS data, all taxa could be confidently delineated to the species and subspecies levels;
2. In addition to comparative mass spectral analysis, data could be obtained on the biological properties of a strain. Thus, we could confidently identify *E. coli*, characterise strain-specific virulence factors and differentiate Shiga-toxin negative from positive strains;
3. This proof of concept study demonstrates that GeLC-MS/MS, which combines the traditional SDS-PAGE platform with LC-MS/MS, has the potential to simultaneously combine strain identification with key pathogenic properties of an isolate and significantly extends the clinical of applications of mass spectrometry in microbiology.

REFERENCES

1. Wibke Kallow, Marcel Erhard, Haroun N. Shah, Emmanuel Raptakis and Martin Welker. MALDI-TOF MS for Microbial Identification: Years of Experimental Development to an Established Protocol. In: '*Mass Spectrometry for Microbial Proteomics'.* Chapter 12. pp. 255- 276. Ed. H.N. Shah and S.E. Gharbia. Wiley, Chichester, 2010.
2. Shah H N, van Steenbergen T J M, Hardie J M, de Graaff J. DNA base composition, DNA-DNA reassociation and isoelectric focusing of proteins of strains designated *Bacteroides oralis. FEMS Microbiology Letters* 1982; **13:** 125-130.
3. Min Fang. Tandem mass spectrometry-based proteomics, protein characterisation and biomarker discovery in microorganisms. In: '*Mass Spectrometry for Microbial Proteomics'.* Chapter 4. pp. 73- 89. Ed. H.N. Shah and S.E. Gharbia. Wiley, Chichester, 2010.