PosterREPRINT

Waters

CLUSTER ANALYSIS OF BACTERIA BASED ON DATA OBTAINED FROM INTACT CELL MATRIX ASSISTED LASER DESORPTION/IONISATION TIME OF FLIGHT MASS SPECTROMETRY

J. J. Bright, H. E. Sutton, V. Edwards-Jones, & D. J. Dare, Manchester Metropolitan University, Manchester, UK.
C. J. Keys, & H. Shah, PHLS, Central Public Health Laboratory, London, UK.
T. McKenna, G. Wells, & M. Lunt, Waters Corporation, Manchester, UK.

Presented at ASM 2002, Salt Lake City, Utah, USA, 19th-23rd May, 2002

Abstract

Intact cell matrix assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS) also known as ICM-MS produces characteristic mass spectral fingerprints of moieties desorbed from the bacterial surface. Mass spectra are obtained within minutes and we have acquired the fingerprints of ~1000 NCTC strains of bacteria in a searchable database. Matching ICM-MS spectra of unknown bacteria to this database provides a rapid new method of identification of bacteria. The relatedness of the mass spectral fingerprints of bacteria in the database can also be determined by cluster analysis of this data. A software package has been written specifically to cluster ICM-MS data. The proximity of one bacterium to another was generated by using a spectral RMS calculation and grouping clusters of bacteria by simple average proximity. In this study the cluster package was tested firstly, to determine the relationship of known bacteria based on their mass spectral fingerprints with their relationship based on more traditional methods and secondly, to determine the level of discrimination that is possible by clustering ICM-MS data.

In an experiment designed to test the cluster analysis package twelve strains of bacteria covering a range of relatedness from Gram-positive to Gram-negative, to two different strains of *Escherichia coli* (both O11:K58(B4):H2) and two strains of *Staphylococcus aureus* (the Type strain and EMRSA-16), demonstrated that clustering was able to resolve the two strains of these species. With a few notable exceptions, the relationship of the strains shown in dendrograms of the ICM-MS data was what might be expected from their accepted relationships e.g. *Escherichia coli* strains were more closely related to each other than *Escherichia hermanii* and more closely related to each other than to *Citrobacter freundii*. However there were some exceptions. *Streptococcus pyogenes* clustered more closely to *Pseudomonas aeruginosa* and all but one of the other Gram-negative bacteria, than to the other Gram-positive bacteria. These results reflect similarities in the moieties desorbed from the surface of these bacteria and therefore, demonstrate relationships that are not always apparent by other methods. The cluster analysis has been applied to groups of bacteria already in the database of ~1000 strains revealing some intriguing results that may help to elucidate the nature of the moieties ionised from the surface of the bacteria. The five O157:H7- VT- strains, of the 28 strains of *Escherichia coli* in the database, all clustered closely together indicating that the O157 antigen may be the major moiety that contributes to the mass spectral fingerprint of these strains. In a similar way, cluster analysis of other groups in the database may give us further clues to the nature of the surface moieties involved in producing ICM-MS spectra.

Introduction

Matrix assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS) of intact bacteria, produces characteristic mass spectral fingerprints of moieties desorbed from the cell [1, 2, 3, 4]. These mass spectra are produced in minutes and can be stored in a database providing the potential for a rapid new method of speciating and typing bacteria. The relationships of different bacterial strains particularly at the level of type, are conventionally determined by cluster analysis. Cluster analysis software was developed by Waters Corporation, UK as part of the MicrobeLynx™ software, specifically for mass spectral fingerprint data from MALDI-TOF-MS. The aim of this study was to evaluate this software and to utilise it to determine the level of reproducibility and discrimination possible by MALDI-TOF-MS.

Waters

Methods

Bacterial strains and Growth conditions

- Twelve NCTC strains of bacteria-6 Gram-positive and 6 Gram-negative were analysed.
- Four batches of 12 replicates of the same strain were included to determine the relative difference in reproducibility.
- Closely and distantly related strains of bacteria were included to determine the level of discrimination.
- NCTC strains, laid down onto Protect beads (Lab M, Bury, Lancs, UK), were revived by streaking a bead onto Columbia Blood (5% v/v) agar (CBA).
- CBA was supplied by the Public Health Laboratory Service accredited laboratories in Chester.
- Incubation: 24 hours at 37°C on CBA in an aerobic atmosphere.
- *Campylobacter jejuni ssp. jejuni* incubation: 24 hours at 37°C on CBA microaerobic (Oxoid Campylobacter System BR 056A gas generating kit).
- Two further sub-cultures were made before MALDI-TOF-MS analysis.

MALDI target preparation

- Bacteria from several colonies applied to target plate wells.
- Four batches of 12 replicates per strain.
- Samples air-dried for at least 1 hour.
- Samples overlaid with 1µL of matrix solution and air-dried.
- Gram-positive bacteria matrix: a saturated solution of 5-chloro-2-mercaptobenzothiazole (Sigma-Aldrich Chemical Company).
- Gram-negative bacteria matrix: a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich).
- Matrix solvent acetonitrile:methanol:water 1:1:1 with 0.1% (v/v) formic acid and 0.01M 18-crown-6 [5].

Instrumentation

- M@LDI-Linear time of flight mass spectrometer (Waters Corporation, Manchester, UK).
- A nitrogen laser giving a 337nm output of 3ns pulse width.
- Laser fluence was set just above the threshold for ion production in the positive ion detection mode.
- Acceleration voltage of +15 kV.
- Mass calibration using average molecular weights from a standard peptide mixture (bradykinin, angiotensin I, glu-fibrinopeptide B, renin substrate tetra decapeptide, ACTH (18-39 clip) all at 1pmol/µL, bovine insulin 2pmol/µL and ubiquitin 10pmol/µL).
- The data acquisition mass range was from m/z 500 to 10000 Da.

Data analysis using Micromass MicrobeLynx™ software

- Spectra were analysed in batches of twelve replicates and compared for reproducibility using the root mean square (RMS) value obtained by comparing each replicate in turn with the average of the other 11 replicates.
- An RMS rejection value of 3 was used to identify outliers significant at the 0.1% level.
- Acceptable spectra were combined to give representative spectra for each strain.
- Dendrograms were produced by an algorithm where the proximity of one organism to another was generated using a spectral RMS calculation.
- Each node of the dendrogram was singly linked.

Waters

- Clusters of organisms were grouped together using a simple average proximity.
- The relative difference axis on the dendrogram represents a relative scale normalised between 0 and 1. A difference of 0 indicates the clusters are exactly the same. A difference of 1 indicates that the clusters are the least similar clusters in the dataset.

Results and Discussion

The reproducibility of MALDI-TOF-MS is illustrated by the clustering of the replicate samples in (**Figure 1**). The cluster analysis software successfully grouped together all the replicate samples of each of the strains with the exception of two *Staphylococcus epidermidis* replicates and one of the EMRSA replicates which were very closely related to the MALDI-TOF-MS spectra of *Staphylococcus aureus* strains (**Figure 1**). The clustering of *Escherichia coli* NCTC 8007 and 8009 into two separate clusters was particularly impressive as these strains both have the same O111:K58(B4):H2 antigens. It can be seen from **Figure 2** that the relative difference between these two strains is significant and about three times that of the replicate variability. This demonstrates the potentially very high level of discrimination obtained by MALDI-TOF-MS. An illustration of the variability/reproducibility of the replicates of *Escherichia coli* NCTC 8007 in terms of the spectra, is given in **Figure 3**. For comparison the spectra of *Escherichia coli* NCTC 8007 and 8009 are given in **Figure 4**.

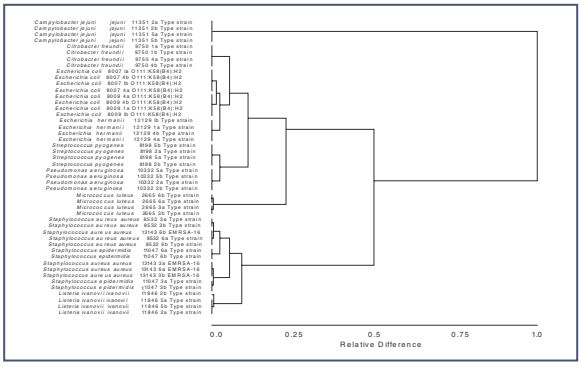

Figure 1. A dendrogram of 4 replicates of 12 NCTC strains of bacteria. The NCTC number, the number of the target plate (1-6), the relative position of the sample on the target plate (a 1st, b 2nd) and antigenic structure follow the species names

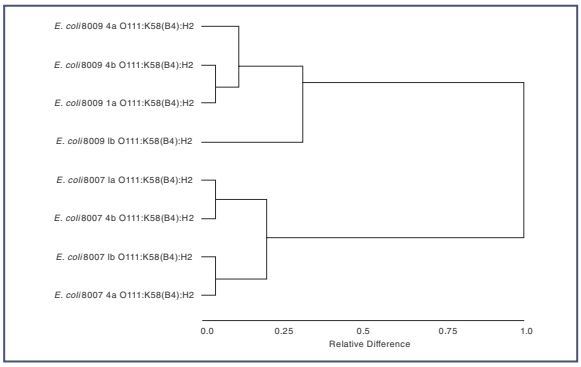

Figure 2. A dendrogram of the MALDI-TOF-MS mass fingerprints of 4 replicates of two Escherichia coli strains with the same O, K, and H antigens. The NCTC number, the number of the target plate (4 or 6), the relative position of the sample on the target plate (a 1st, b 2nd) and antigenic structure follow the species names

Waters

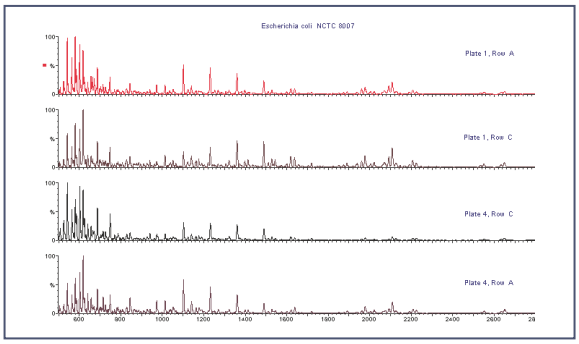

Figure 3. MALDI-TOF-MS spectra of 4 replicates of Escherichia coli NCTC 8007

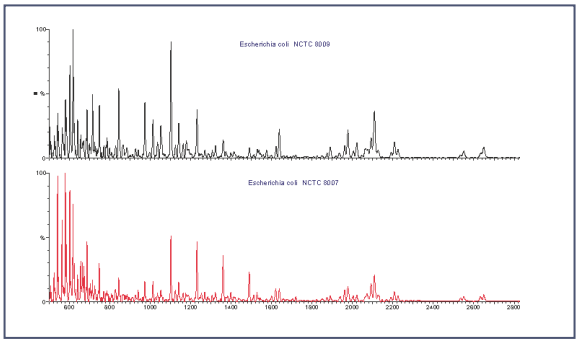

Figure 4. MALDI-TOF-MS spectra of Escherichia coli NCTC 8007 and Escherichia coli NCTC 8009

In general, the pattern of clustering of the different species based on their mass fingerprints illustrated in the dendrogram in **Figure 1**, follows what might be expected from our general knowledge of the relationships of the strains used in this study. For example, the two *Escherichia coli* strains clustered very closely together followed by *Escherichia hermanii* then *Citrobacter freundii* and the Gram-positive and Gram-negative strains largely clustered together. However, there are some exceptions and in particular *Campylobacter jejuni* was twice as different as the Gram-positive and largely Gram-negative clusters. It can be seen in **Figure 5** that this difference can be accounted for by the large broad peaks at ~m/z 5700-6000 for *Campylobacter jejuni*, which are not present for *Citrobacter freundii* or, any of the other strains analysed in this study. Other surprising relationships were observed for *Streptococcus pyogenes* which clustered very close to *Pseudomonas aeruginosa* and the other Gram-positive bacteria. *Micrococcus luteus* was more closely related to the Gram-negative cluster than the largely Gram-positive cluster. These relationships reflect the similarities that these bacteria have in the moieties that are ionised by MALDI-TOF-MS.

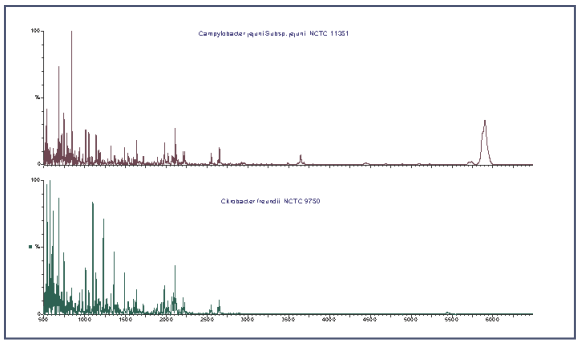

Figure 5. MALDI-TOF-MS spectra of Campylobacter jejuni NCTC 11351 and Citrobacter freundii NCTC 9750

Given the success of the cluster analysis algorithm, it was applied to a mass fingerprint database of 28 NCTC *Escherichia coli* strains. The resulting dendrogram clustered closely together all five O157 VT- strains (**Figure 6**). This clustering is probably a result of similar mass ions being produced by the O157 antigen as has been suggested previously in a study by Bright *et al.*, 2002 [6]. The only other O157 strain in the database failing to cluster with the O157 VT- strains was NCTC 10964 (O157:K88a,c:H19). The presence of a K antigen may have had an influence on the mass spectral profile of this strain. Other evidence that support these suggestions are that two strains with the same O and K antigens (NCTC 8007 O111:K58(B4):H2 and NCTC 8009 O111:K58(B4):H2), both clustered together whilst, two pairs of strains with the same O antigen but different K antigens (NCTC 9020 (O20:K17(L17):H-) NCTC 10864 (O20:K61:H-) and NCTC 9002 O2:K?:H4, NCTC 11151 O2:K1:H4), were not closely related in the dendrogram (**Figure 6**).

Waters

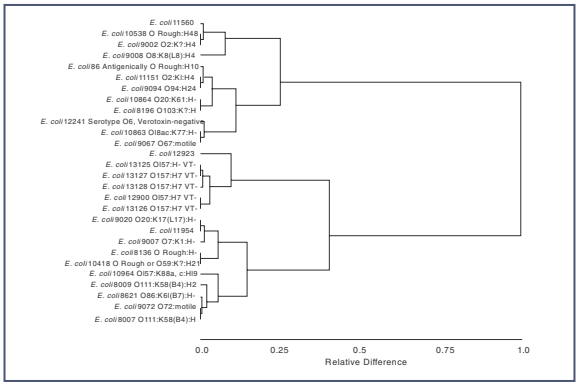

Figure 6. A dendrogram of the MALDI-TOF-MS mass fingerprints of 28 NCTC strains of Escherichia coli. The NCTC number and antigenic structure follow the species names

Conclusion

- Clustering of replicate MALDI-TOF-MS mass fingerprints of a range of bacteria demonstrates the reproducibility of the technique.
- The discrimination of the MALDI-TOF-MS technique is very high.
- *Escherichia coli* strains with the same O111:K58(B4):H2 antigens are differentiated by MALDI-TOF-MS.
- The pattern of clustering of MALDI-TOF-MS data of bacteria generally reflects the known relationships of the bacteria to each other.
- Clustering of MALDI-TOF-MS data revealed some unexpected relationships:
  - The distant relationship of *Campylobacter jejuni* to the other strains was due to the presence of large broad peaks at m/z ~5700-6000.
  - The close relationships of some Gram-positive bacteria with Gram-negative bacteria probably reflects, the ionisation of shared moieties.
- Mass fingerprints of *Escherichia coli* strains with the same O antigen (e.g. O157) cluster together but this relationship can be influenced by the presence of a K antigen.
- MALDI-TOF-MS has the potential to be a very rapid, discriminating, method of typing bacteria.

References

1 Claydon, M. A., Davey, S. N., Edwards-Jones, V., Gordon, D. B., 1996. The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotech.* **14**, 1584-1586.

2 Holland, R. D., Wilkes, J. G., Rafii, F., Sutherland, J. B., Persons, C. C., Voorhees, K. J., Lay, J. O., 1996. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption with time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **10**, 1227-1232.

3 Krishnamurthy, T., Ross, P. L., Rajamani. U., 1996. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **10**, 883-888.

4 Welham, K. J., Domin, M. A., Scannell, D. E., Cohen, E., Ashton, D. S., 1998. The characterisation of microorganisms by matrix-assisted laser desorption ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **12**, 176-180.

5 Evason, D. J., Claydon, M. A., Gordon, D. B. 2000. Effects of ion mode and matrix additives in the identification of bacteria by intact cell mass spectrometry. *Rapid Commun. Mass Spectrom.* **14**, 669-672.

6 Bright, J. J., Claydon, M. A., Soufian, M.A., Gordon, D. B., 2002. Rapid Typing of Bacteria Using Matrix Assisted Laser Desorption Ionisation Time of Flight Mass Spectrometry and Pattern Recognition Software. *J. Microbiol. Methods*, **48**, 127-138.

PosterREPRINT

Author to whom all correspondence should be addressed:
Thérèse McKenna
Waters Corporation
(Micromass UK Limited)
Floats Road, Wythenshawe
Manchester, M23 9LZ
Tel: + 44 (0) 161 946 2400
Fax: + 44 (0) 161 946 2480
e-mail: therese.mckenna@micromass.co.uk

WATERS CORPORATION
34 Maple St.
Milford, MA 01757 U.S.A.
T: 508 478 2000
F: 508 872 1990
www.waters.com

Made in the United Kingdom

Waters
RIGHT ON TIME.

MICROMASS®
MS TECHNOLOGIES

Certificate No: 951387