Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Surface Science 252 (2006) 6719–6722
applied surface science
www.elsevier.com/locate/apsusc

ToF-SIMS studies of *Bacillus* using multivariate analysis with possible identification and taxonomic applications

C.E. Thompson [a], J. Ellis [b], J.S. Fletcher [a], R. Goodacre [b], A. Henderson [a], N.P. Lockyer [a,*], J.C. Vickerman [a]

[a] *Surface Analysis Research Centre, School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M60 1QD, UK*
[b] *School of Chemistry, University of Manchester, Manchester M60 1QD, UK*

Received 12 September 2005; accepted 15 February 2006
Available online 22 May 2006

Abstract

In this paper we discuss the application of ToF-SIMS with an Au_3^+ primary ion beam, combined with principal components analysis (PCA) and discriminant function analysis (DFA) for the identification of individual strains of two *Bacillus* species. The ToF-SIMS PC-DFA methodology is capable of distinguishing bacteria at the strain level based on analysis of surface chemical species. By classifying the data using hierarchical cluster analysis (HCA) we are able to show quantitative separation of species and of these strains. This has taxonomic implications in the areas of rapid identification of pathogenic microbes isolated from the clinic, food and environment.

Keywords: ToF-SIMS; *Bacillus*; Multivariate analysis; Taxonomy; PCA; DFA

1. Introduction

Bacillus cereus is described as one of the most taxonomically confusing areas of the bacilli [3]. This has significance in today's modern medicine (antibiotic resistance), food industry (for food spoilage organisms) and bioterrorism (*B. anthracis* which is responsible for anthrax and is in the same taxonomic division as *B. cereus*) where bacterial diagnostic is an important tool, especially where minute biochemical differences between organisms are difficult to detect. Conventional taxonomic methods and whole cell lipid profiling have been shown already to be able to discriminate between *B. cereus* and *B. subtilis* [4], and so this was considered a good model system for SIMS.

The need for rapid, reproducible and accurate analyses for the detection and identification of bacteria has long been a target for research. The family *Bacillus* has been divided into three main divisions of which *B. cereus* and *B. subtilis* fall within the same division, they are closely related yet distinct by DNA–DNA homology and physiological traits [4]. *B. cereus* and *B. subtilis* are known to be closely related [5] and so morphological methods, such as classical culturing and microscopy may not readily discriminate between species. For bacterial taxonomic relationships to be examined from chemical data cluster analysis both ordination plots (principal components (PCA) or discriminant function (DFA) analyses) and dendrograms (tree like figures generated from hierarchal cluster analysis (HCA)) are typically employed. In these methods, distances between organism can be used to assess their relative similarity to one another.

The benefit of chemometric analysis of mass spectrometric data has been highlighted previously showing multivariate analysis as a powerful tool for raw data sets [2,6,7]. PCA has previously been used with ToF-SIMS spectra to show differences between the spores and vegetative cells of *Bacillus megaterium* and four different strains of yeast [1,8]. In other surface-based analysis of bacteria that use biochemical profiling, multivariate analyses such as PCA and DFA has allowed for accurate comparison of ATR–FT-IR spectra from clinical isolates of *B. cereus* and *B. subtilis* [9]. MALDI-ToF-MS protein profiling has been used as a taxonomic tool for differentiating spores of *Bacillus* [10], although this technique is somewhat less surface-specific.

In the present study we assess the ability of a combined ToF-SIMS and multivariate analysis to distinguish two Gram

* Corresponding author.
E-mail address: nick.lockyer@manchester.ac.uk (N.P. Lockyer).

0169-4332/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.apsusc.2006.02.237

Table 1
Bacterial species and the corresponding strain number

Species	ID number	Dendrogram identifier	Plot strain identifier
B. cereus	B0002^T	Ca	(●)
	B0550	Cb	(○)
	B0702	Cc	(⊘)
B. subtilis	B1382	Sa	(▲)
	B0098	Sb	(▼)
	B0014^T	Sc	(△)

^{T} indicates the type strain.

positive *Bacillus* species (*B. cereus* and *B. subtilis*) comprising six individual strains (see Table 1). Supervised clustering methods such as DFA require prior knowledge of groups in the data whilst unsupervised methods like PCA use no prior knowledge of groups structure. Therefore, as a comparison both supervised and unsupervised results will be shown. Finally, dendrograms were produced from PC-DFA space as they describe the inherent hierarchical structure of class division within the data set.

2. Experimental

2.1. Sample preparation

Table 1 shows the species and strains analysed: further details of these organisms are available in Vaidyanathan et al. [11]. Vegetative cells from axenic cultures were grown on Blood M agar (without blood) for 12 h on separate plates at 37 °C. This process was repeated over 5 d to produce five biological repeats. The bacteria were washed from the plates and suspended in 1 ml of sterile distilled water and stored at −80 °C until needed.

Aliquots of sample were placed onto single 5 mm × 5 mm silicon wafers mounted onto copper sample stubs. ToF-SIMS analysis was performed on a BioToF-SIMS instrument using a 15 keV Au_3^+ primary ion beam (Ionoptika, UK) with a total primary ion dose density 4×10^{10} ions cm^{-2}. A total of 135 mass spectra were taken (three 'machine' repeats for each 'biological' repeat).

2.2. Data analysis

The ToF-SIMS spectra were binned to 1 mass unit prior to chemometric analysis to reduce the computational power required for the analysis. Although binning the data in this manner may reduce the mass resolution of the data the instrument was not run in maximum mass resolution mode and in this case these effects are considered negligible. The remaining files still contain a characteristic fingerprint from each of the samples. The mass range *m/z* 51–1900 was used in order to include possible characteristic lipid species and exclude salts that vary extensively from one sample to another. Data were sum normalised by the total ion intensity within the spectrum to a total of 100% and mean-centered, to remove spectral variation in terms of peak intensities and noise. Matlab (MathWorks, Inc., MA, USA) was used to analyse the data set by performing PCA using the NIPALS algorithm, reducing data dimensionality and preserving variance [9].

The normalized PCA data values were used for DFA and the data were allocated groups for the species and strain they represent (two species and six strains). HCA was programmed to classify data by first generating a similarity matrix from ordinary Euclidean distances between samples in DFA space, and an agglomerative algorithm was used to produce a dendrogram of relationship within these spectra.

3. Results

PCA of the complete data set in the mass range *m/z* 51–1900 resulted in intermixing of species data and some intermixing and no tight clustering of strains (Fig. 1(a)). When examining the strain level separation the PCA resulted in some degree of separation, the majority of the data showed significant overlap making it difficult to see clear clustering. As the names of each sample in the test set are known, it is possible to use the supervised DFA technique to facilitate groupings based on

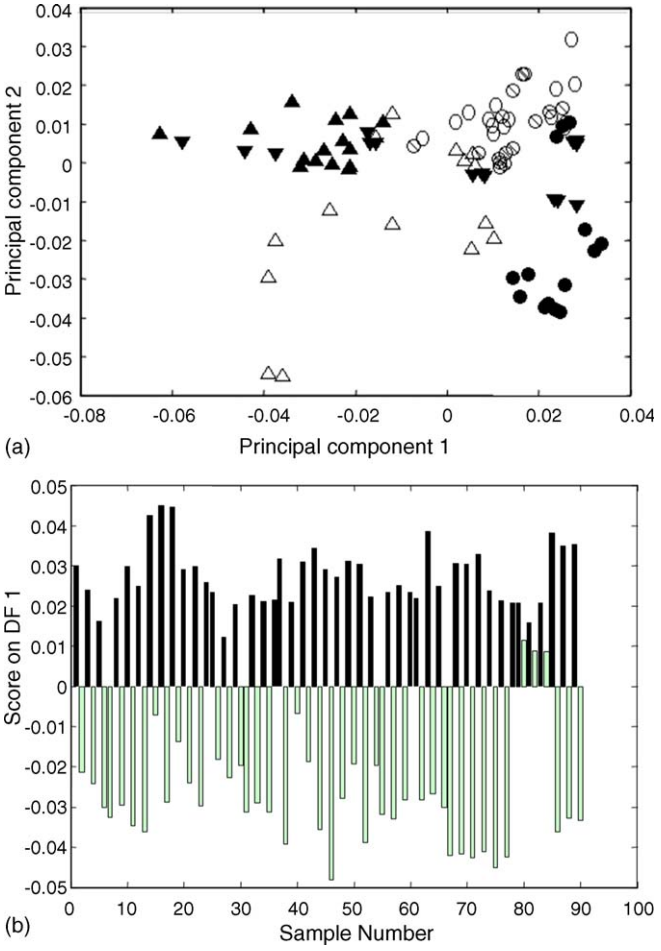

Fig. 1. (a) 2D PCA plot of ToF-SIMS data from *Bacillus cereus* and *Bacillus subtilis* strains (PC 1 and PC 2 exhibit 35.50% and 19.24% of the total variance). (b) A bar chart showing how species are separated within the DF 1 space (black bars represent *B. cereus*, grey bars represent *B. subtilis*).

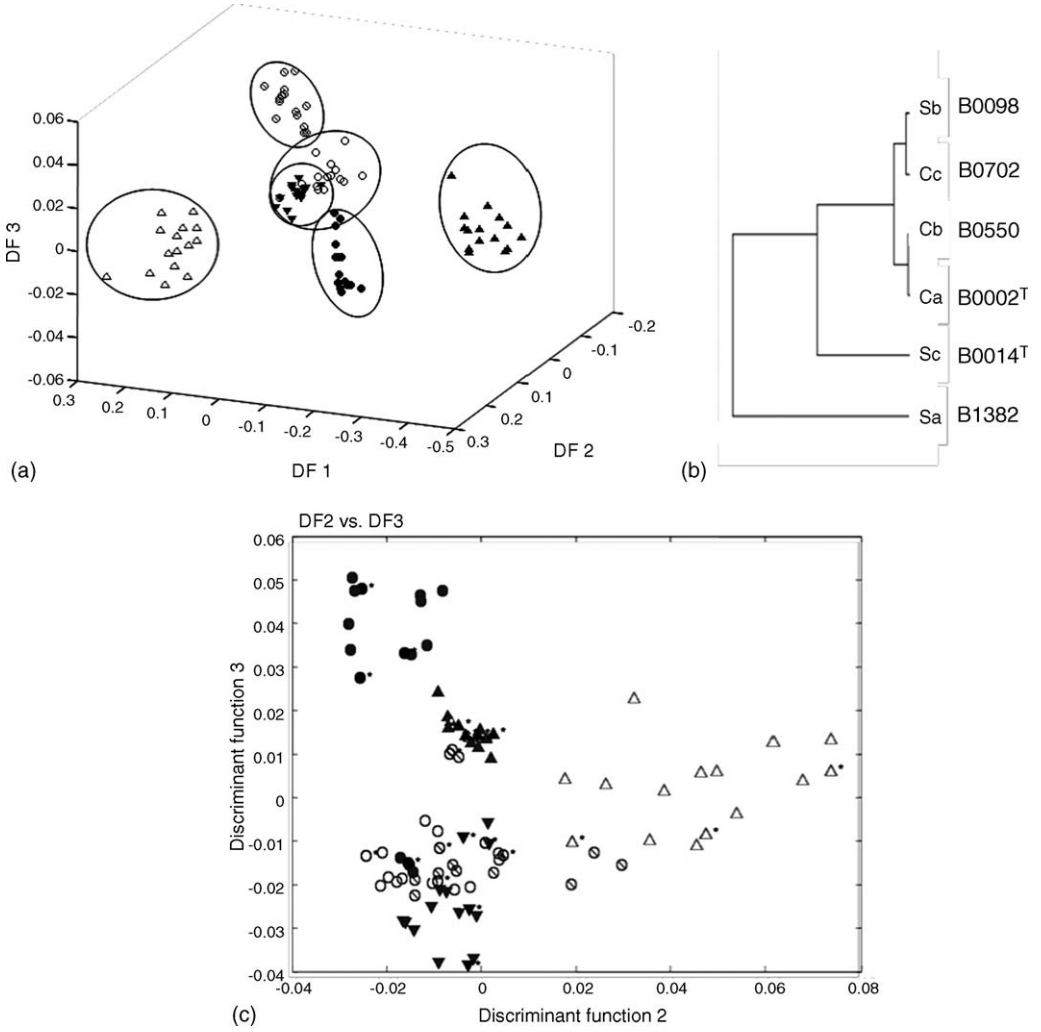

Fig. 2. (a) DFA plot of strain groups (note circles are included for visual clarity) with the percentage variance of DF 1 (35.42%), DF 2 (19.37%) and DF 3 (6.35%). (b) HCA dendrogram showing the average data separated on the distance between one strain and the next. Refer to Table 1 for dendrogram strain identifiers. (c) DF2 against DF3 plot showing test set data projection (test set data accompanied by an asterisk).

similarities between the strains. The discriminating power of the combined ToF-SIMS and multivariate analysis is observed in the plots shown.

Fig. 1(b) shows sample number separation in the DF1 space, clearly showing the data separated into the two species groups. Fig. 2(a) shows the result of applying *a priori* knowledge to the data, in this case groupings for species (*B. cereus* and *B. subtilis*). A test set of data was projected into the PC-DFA space of the training set in order to validate the clustering. The test set showed successful clustering (Fig. 2(c)).

Fig. 2(a) shows the pseudo-3D DFA plot for DF 1, DF 2 and DF 3. There are two clear clusters for *B. subtilis* strains B1382 and B0014, yet the third *B. subtilis* strain (B0098) has clustered with the *B. cereus* strains. DF 1 therefore discriminates between *B. subtilis* B1382 and B0014 (type strain). Clearly DF 3 separates the strains of *B. cereus* (circle symbols). There appears to be a defined separation of all six strains, with a degree of overlap between *B. cereus* strain B0550 and *B. subtilis* strain B0098. This overlap is an indication of surface similarity between B0550 and B0098. It is necessary to achieve good separation to perform a hierarchical cluster analysis.

The HCA (Fig. 2(b)) produced a dendrogram showing the relationship between these strains. This shows the resulting natural relationships (those that are from the biochemistry of the organism and not a result of machine differences as these should be removed with data preprocessing) detected by the ToF-SIMS analysis. This taxonomic classification allows one to observe the relationship between the data in a few simple steps, something that may be difficult without the use of multivariate analysis. It is clear that all the chosen strains separate and that strain *B. subtilis* (strain B1382) is the most dissimilar strain on the basis of our analysis. Not only has it separated from the *B. cereus* but also has not shown a close relationship with the other *B. subtilis* strains. However there is clear similarity between *B. subtilis* B0098 and *B. cereus* B0550 as seen in the previous DFA plot (Fig. 2(a)) The *B. cereus* have grouped closer together indicating that their surface chemistry has a greater similarity than the *B. subtilis* strains have to one another, within their genotype.

4. Discussion

The bacterial cell surface, with polypeptides, glycoproteins, polysaccharides and phospholipids, is very complex and contains a wealth of biochemical information. With this in mind we believe that ToF-SIMS is an ideal method for probing the bacterial cell surface and might yield useful discriminatory information.

Detailed surface analysis, such as afforded by ToF-SIMS allows greater knowledge of the biochemistry within that region. Winder and Goodacre [9] found that FT-IR and ATR–IR was able to separate strains at the subspecies level, ATR–IR being a surface specific analysis technique. The surface of bacteria allow for taxonomy based on the variations between species in terms of their chemistry.

5. Conclusion

The use of multivariate analysis makes ToF-SIMS a successful technique for the discrimination of bacterial cells by their surface differences. Not only can the data be examined in terms of biochemistry from PCA and DFA loadings plots (not shown), but the production of dendrograms from HCA allows relationships between the bacterial strains to be elucidated. This information may then be used to classify organisms of unknown type or origin. Further work will examine the biochemical strain differences in more detail.

Acknowledgments

The authors wish to thank BBSRC and EPSRC for funding.

References

[1] C.E. Thompson, H. Jungnickel, N.P. Lockyer, G.M. Stephens, J.C. Vickerman, Appl. Surf. Sci. 231/232 (2004) 420.
[2] S.J. Pachuta, Appl. Surf. Sci. 231/232 (2004) 217.
[3] K.K. Hill, L.O. Ticknor, R.T. Okinaka, M. Asay, H. Blair, K.A. Bliss, M. Laker, P.E. Pardington, A.P. Richardson, M. Tonks, D.J. Beecher, J.D. Kemp, A. Kolstø, A.C. Lee Wong, P. Keim, P.J. Jackson, Appl. Environ. Microbiol. 70 (2) (2004) 1068.
[4] G.E. Black, A.P. Snyder, K.S. Heroux, J. Microbiol. Meth. 28 (1997) 187.
[5] S.G. Bavykin, Y.P. Lysov, V. Zakhariev, J.J. Kelly, J. Jackman, D.A. Stahl, A. Cherni, J. Clin. Microbiol. 42 (8) (2004) 3711.
[6] S.L. McArthur, M.S. Wagner, P.G. Hartley, K.M. McLean, H.J. Griesser, D.G. Castner, Surf. Int. Anal. 33 (12) (2002) 924.
[7] O.D. Sanni, M.S. Wagner, D. Briggs, D.G. Castner, J.C. Vickerman, Surf. Int. Anal. 33 (9) (2002) 715.
[8] H. Jungnickel, E.A. Jones, N.P. Lockyer, S.G. Oliver, G.M. Stevens, J.C. Vickerman, Anal. Chem. 77 (2005) 1740.
[9] C.L. Winder, R. Goodacre, Analyst 129 (2004) 1118.
[10] D.N. Dickinson, M.T. La Duc, M. Satomi, J.D. Winefordner, D.H. Powell, K. Venkateswaran, J. Micro. Meth. 58 (2004) 1.
[11] S. Vaidyanathan, J.J. Rowland, D.B. Kell, R. Goodacre, Anal. Chem. 73 (2001) 4134.