Rapid Biomarker Profiling Of Escherichia coli Utilising MALDI-TOF Mass Spectrometry and a Multivariate Pattern Recognition Approach

Remi Momo*†[1], Elaine Martin Obe[2], Gary Montague[3], Ronan O'kennedy[4], Jane Povey[5], Chris O'malley[6], and Mark Smales[7]

[1]School of Chemical Engineering and Advanced Materials – Newcastle University – Newcastle upon Tyne, United Kingdom
[2]School of Chemical Engineering and Advanced Materials – Newcastle University – Newcastle upon Tyne, United Kingdom
[3]School of Chemical Engineering and Advanced Materials – Newcastle University – Newcastle upon Tyne, United Kingdom
[4]School of Chemical Engineering and Advanced Materials – Newcastle University – Newcastle upon Tyne, United Kingdom
[5]School of Biosciences – University of Kent – Canterbury, Kent, UK, United Kingdom
[6]School of Chemical Engineering and Advanced Materials – Newcastle University – Newcastle upon Tyne, United Kingdom
[7]School of Biosciences – University of Kent – Canterbury, Kent, UK, United Kingdom

Abstract

Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) has been exploited extensively in the field of microbiology for the investigation of bacterial species, the detection of biomarkers and early disease diagnosis. One of the key issues involved in MALDI-TOF MS studies is the need to reduce the high-dimensionality of the data generated. In the context of biomarker detection, analysis of such complex spectra data may cause variances pertaining to the experimental hypothesis to be confounded with variations from experimental sources. More so, one of the challenges is to detect the biology-based variability related to the samples being analysed to facilitate the identification of specific protein biomarker ions involved. In the study reported, chemometric data analysis techniques were applied to MALDI-TOF MS data obtained from E. coli cell samples. The technique of primary interest was a multivariate pattern recognition technique, Partial Least Squares-Discriminant Analysis (PLS-DA). It was applied to investigate protein biomarkers relating to the age of E. coli cells grown in liquid culture. After an initial pre-processing step, the full dataset of 366 samples was analysed by Principal Component Analysis (PCA) to get an overview of the data. There was strong evidence of separation between the samples of cells at mid-log exponential phase culture from the remaining two groups of samples (stationary and prolonged stationary phases), along the second principal component. The samples were then modelled by PLS-DA in order to classify them based on the growth phases of the cells. The mass-to-charge ratio (m/z) biomarker peaks responsible for their classification were also

*Speaker
†Corresponding author: r.a.m.momo@newcastle.ac.uk

identified. Results suggest that exponential phase cultures could be distinguished from the rest of the cell with ions near the m/z 5095, 6254, 6315, 6410, 7273, 7871, 9535, 10694, 11199 and 12223. A Swiss-Prot/TrEMBL database identified and attributed most m/z ion signals from exponential phase cultures to series of ribosomal proteins (L22, L23, L25, L29, L31, L32, L33, S20 and S22); those for stationary phase cultures attributed to nucleoid-associated proteins (MccB17, IhfA and Dps); and toxin-antitoxin (TA) module products (ChpS antitoxins, YefM, RelE and RelB) for death phase cultures which are stress-response molecules. These results are not surprising since up to 45% of the mass of rapidly growing E. coli cells correspond to ribosomes, and up to 21% of the cell's protein content is ribosomal. This suggest that cells probably grow faster at early growth phases and assemble more ribosomes in order to produce stress-response elements that help cells survive unfavourable growth conditions at later growth phases when nutrients become depleted. The recognition of these protein m/z signals suggest that these proteins can be used as growth-phase-associated biomarkers to distinguish between E. coli cultures at different growth phases.

Keywords: Biomarker, Escherichia coli, MALDI, TOF Mass Spectrometry, Multivariate pattern recognition