Does LC Separation of Intact Proteins Hinder Identification of Bacterial Markers?

Melinda A. McFarland ; Denis Andrzejewski; John H. Callahan; Steven M. Musser
FDA-CFSAN, College Park, MD

Novel Aspect: Comparison of detected intact masses, top-down, and bottom-up for identification of strain specific bacterial markers.

Introduction

Intact protein expression profiling and subsequent identification of marker proteins is a powerful tool for bacterial strain differentiation. Advances in protein chromatography and deconvolution software make it possible to generate reproducible accurate mass – retention time profiles of > 500 features per strain. Advances in top-down technology and informatics make it possible to identify > a hundred proteins per sample by online LC-MS/MS, while maintaining intact mass for identification of SNP based biomarkers. However, protein level LCMS biases against classes and sizes of proteins and bottom-up proteomics identifies ten times the number of proteins. We investigate limits to mapping intact masses to bottom-up protein identifications, SNP identification by top-down versus bottom-up, and which proteins are lost with protein chromatography.

Methods

Salmonella proteins are extracted with a Barocycler (Pressure Biosciences) and separated with a Prosphere P-HR column (150mm x 2.1mm). LCMS was performed on a QTOF Premier (Waters) and deconvoluted to accurate mass, retention time, and intensity profiles with ProteinTrawler6 (BioAnalyte). Top-down and bottom-up MS/MS are performed on an Orbitrap (ThermoFisher) using a TriVersa Nanomate source (Advion Biosciences). LCMS fractions were infused for top-down identification or digested and analyzed by LC-MS/MS. On-line top-down LC-MS/MS was also performed using the same LC gradient to match retention times. The identical lysate was run on a 1D SDS-PAGE gel, digested and analyzed by LC-MS/MS. Bottom-up spectra were searched with the Mascot algorithm and top-down spectra were searched with the ProSightPC algorithm.

Preliminary Data

LCMS data show good protein level separation and reproducible extraction of protein molecular weight and retention time profiles, as well as the presence of type specific bacterial marker masses. The one hundred most abundant proteins from intact protein profiles have been identified by on-line and off-line top-down MS/MS. Bottom-up digestion of the corresponding LC fractions yields significantly more identifications, but database deficiencies mean those identifications can not always be mapped back to the intact mass. This can complicate definitive identification of differential protein profile marker masses. However, bottom-up analysis of fractions has the benefit of identifying proteins that may be chromatographically separated but not detected or resolved by intact protein LCMS. Both of these experiments limit bacterial strain analysis to the proteins that are amenable to intact protein chromatography. GeLC analysis from SDS-PAGE digestions yielded many additional identifications from the bottom-up, most of which were in a mass range not amenable to top-down MS/MS, were present at an abundance that is too low for intact protein mass spectrometry, or did not separate on the column. However, while SDS-PAGE maintains an approximate link to the intact protein mass, the comparatively small mass shifts induced by SNP differences between different strains can not be resolved by the gel. Without an intact mass measurement, we are dependent on the database and peptide sequence coverage for detection of the presence of SNPs. Considering the diversity of bacteria, relatively few strain specific bacterial proteins are found in protein sequence databases. On the other hand, the mass range and total number of proteins identified by GeLC analysis surpasses the number of proteins amenable to intact protein HPLC separation. We will also investigate the difficulty in referencing proteins identified by searching bottom-up data against incompletely annotated databases back to an intact mass.