CLASSIFICATION AND IDENTIFICATION OF MICROBES USING A MASS SPECTROMETRY PROTEOMIC METHOD (MSPM)

Samir V. Deshpande[1]*, Rabih E. Jabbour[2], Michael Stanford[3], Charles Wick[3], Alan Zulich[3]

[1]Science & Technology Corporation, Edgewood, MD 21040;

[2]SAIC, Inc., Aberdeen Proving Ground, MD 21010;

[3]US Army Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD 21010

* Email:samir.deshpande@us.army.mil

Abstract:

We have developed suite of bioinformatics algorithms for automated identification and classification of microbes based on comparative analysis of protein sequences. This application uses sequence information of microbial proteins revealed by mass spectrometry-based proteomics for rapid bacterial identification (ABOid). This algorithms transforms results of searching product ion spectra of peptide ions against a protein database, performed by commercially available software (e.g. SEQUEST), into a taxonomically meaningful and easy to interpret output. To achieve this goal we constructed a custom protein database composed of theoretical proteomes derived from all fully sequenced bacterial genomes (more than 1167 microorganisms) in a FASTA format. Each protein sequence in this database is supplemented with information on a source organism and chromosomal position of each protein coding open reading frame embedded into a header. In addition this information is linked with a taxonomic position of each database bacterium.

ABOid analyzes SEQUEST search results files to provide the probabilities that peptide sequence assignments to a product ion mass spectrum (MS/MS) are correct and uses the accepted spectrum–to-sequence matches to generate a sequence-to-bacterium (STB) matrix of assignments. Because peptide sequences are differentially present or absent in various strains being compared this allows for the classification of bacterial species in a high throughput manner. For this purpose, STB matrices of assignments, viewed as assignment bitmaps, are next analyzed by a ABOid module that uses phylogenetic relationships between bacterial species as a part of decision tree process, and by applying multivariate statistical techniques (principal component and cluster analysis), to reveal relationship of the analyzed unknown sample to the database microorganisms. Our bacterial classification and identification algorithm uses assignments of an analyzed organism to taxonomic groups based on an organized scheme that begins at the phylum level and follows through classes, orders, families and genus down to strain level.

In summary, ABOid is capable of revealing the identification of microbial mixture contents. ABOid does not require prior knowledge of the sample and can be applied to pure cultures and mixtures. ABOid allows for strain identification based on comparative analysis of protein sequences. Bacterial strains not in database were identified to their close-neighbor species and Statistical and visualization tools of the ABOid allow its utilization by non-specialists.