The development of the method for the analysis of proteomics data

Dmitry S. Ischenko, V. Popkov
Moscow Institute of Physics and Technology, Dolgoprudny, Russia, ischenko.dmitry@gmail.com
Dmitry G. Alexeev
Scientific-Research Institute of Physicochemical medicine, Moscow, Russia, exappeal@gmail.com

In the analysis of mass spectrometry data of unannotated sample the problem is in an adequate evaluation of the results. The question is: what to choose as a protein database for searching? There are several solutions: One of them is to take one of the samples studied and annotated, the second is to create an aggregate database of all known genes, or, if the sequenced data exists, create your own database (taking into account the single nucleotide polymorphisms).

The work on the study of the reliability of the resulting search for samples of proteins *Helicobacter pylori, Escherichia coli, Bacillus subtilis, Staphylococcus aureus, Deinococcus maricopensis.* Searches were carried out on various protein databases: taken separately genomes, database of all known genes and databases using MSMSPdbb (Multi-Strain Mass Spectrometry Prokaryotic DataBase Builder) [1]. Results were compared with the authentic proteins produced during sequencing samples and comparing it with annotated strains.

We assessed the quality and reliability of the detection of peptides and proteins in searching by various databases. We propose methods to improve the accuracy of peptides and proteins detection and show the dependence of the actual detection and the distance between strains., take into account features of peptides, such as their uniqueness and conservatism. We made assumptions about the preferred protein databases for the search and several ways to increase detection accuracy. The analysis was conducted using the application Mascot v.2.2.07[2], Protein Pilot Software 4.5, Oracle 11g database, R language.

1. Gustavo A. de Souza, Magnus O. Arntzen (2010) MSMSpdbb: providing protein databases of closely related organisms to improve proteomic characterization of prokaryotic microbes. Bioinformatics, 26(5):698-9

2. Koenig T, Menze BH, Kirchner M, et al. (2008). "Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics". J. Proteome Res. 7 (9): 3708–17.