

The Golm Metabolome Database between Metabolite Substructure Prediction and Reporting of Metabolomic Data Sets for the Scientific Community

Joachim Kopka¹, Jan Hummel¹, Nadine Strehmel² and Dirk Walter¹

¹Max Planck Institute of Molecular Plant Physiology (MPIMP), Germany

²Leibniz Institute of Plant Biochemistry (IPB), Department of Stress and Developmental Biology, Germany

Gas chromatography coupled to mass spectrometry (GC-MS) is a widely applied routine technology for the large scale screening and discovery of novel metabolic biomarkers. However, the majority of mass spectral tags (MSTs) recorded by this and other mass spectrometry-based metabolomic methods remains unidentified largely due to the lack of authenticated pure reference substances, which are required for unambiguous compound identification. We accessed the information on identified reference compounds stored in the Golm Metabolome Database (GMD) and applied supervised machine learning approaches for the classification of yet unidentified MSTs. We thereby extended conventional mass spectral and retention index library searches towards a more refined assessment of yet unknown MSTs. Structural feature extraction was applied to sub-divide the metabolite space contained in the GMD and to define frequently occurring substructures of metabolites which were suitable for our cheminformatic approach. Decision tree (DT)-based prediction of the most frequent substructures was performed using mass spectral features and RI information. We established a highly sensitive and specific analysis of biologically relevant sub-structures contained within the GMD compendium which returns potential substructures for yet unknown MSTs from GC-MS profiles together with probability criteria and the rules that led to the respective prediction. The underlying set of DTs can be inspected by the user and are made available via a new user friendly web-based interface to GMD and via SOAP (Simple Object Access Protocol)-based web services. Thus the grounds are prepared for the public reporting of metabolomic data sets for the scientific community covering both the known and the yet non-identified metabolites.

If this proposed topic should not be the one you intended please let me know, I can easily come up with a less generic talk than this on systems application with focus either on yeast temperature stress physiology, cyanobacterial profiling and stable isotope probing to understand photorespiration, or plant environmental stress physiology. Because Metabolomics-2012 appears to foster the interactions between basic science/biology and human/ clinical systems biology the above generic topic centering on our data base development appears to be the best choice.

Biography

Dr. Joachim Kopka is Research Group Leader at the Max-Planck-Institute of Molecular Plant Physiology (MPIMP), Germany and leading the Applied Metabolome Analysis Science Infrastructure group at the MPIMP, Potsdam-Golm, Germany. He did his PhD from Westfälische Wilhelms Universität, Münster, Germany. He also served as Leader of the technical center Bioanalytics, Metanomics, Berlin, Germany, Leader of the technical center Bioinformatics/ Datamining, Metanomics, Berlin, Germany. He is co-founder of the Metanomics GmbH & Co. KGaA company (Metanomics), Berlin, Germany. He has more than 90 publications and 14 Editorial articles in various journals of national and international repute. He has 4 patents granted. He organized many international conferences and workshops.