Towards a complete automatisation of non-target screening approaches with high resolution mass spectrometry

SASCHA LEGE, CHRISTIAN ZWIENER

University of Tübingen, Environmental Analytical Chemistry, Hölderlinstrasse 12, 72074 Tübingen, Germany

New generations of mass spectrometers combine high sensitivity in the picomolar to femtomolar range with high mass resolving power. Therefore, they are more and more applied for the analysis of complex samples (e.g. surface water) and they might replace in the future low-resolution instruments that are nowadays usually used for routine analysis and quantification purposes. The benefit from highly resolved mass spectra is the versatility of the data evaluation possibilities, because the analyst is not only restricted to the investigation of a set of known compounds. Instead, by applying non-target screening approaches, information is already acquired for all those compounds that can be analyzed with the applied technique. The analytes can be subsequently identified based on accurately measured masses, isotope distributions, and high resolution MS^n spectra.

Due to the development of autosamplers, it is generally possible to perform such non-target screening analyses in an unsupervised and high-throughput manner. However, the amount and complexity of data generated by these approaches is so high that sophisticated software solutions are required for their evaluation. Usually, the tools for data processing are supplied by the mass spectrometer vendors and univariate, as well as multivariate statistical evaluation techniques are included. By this, it is possible to reduce the complexity of the datasets and to end up only with significant analytes. However, these tools have all in common that they have to be operated manually by the analyst after sample measurement. Therefore sample throughput decreases when measurements finish for example at night or on weekends.

Here, we present a potential solution for this issue which is based on a combination of software tools from Agilent Technologies and the freeware KNIME, a graphical workbench that was first released in 2006. KNIME is increasingly applied to handle “Big Data“ in scientific research, because the only limitation for the data size is the availability of hard drive space rather than the random access memory. Additionally, KNIME benefits from an easy extensibility, because selected chem- and bioinformatic tools (e.g. OpenMS) are already available as plug-ins in KNIME or own tools, developed in Java, Perl, Python, or R, can be incorporated into the evaluation process. Our workflow was applied for a non-target screening of effluent samples from 22 wastewater treatment plants. Overall, 85 measurements were performed and analyte information were immediately extracted with the MassHunter Qualitative Analysis (Agilent Technologies) tool. The exported feature information was subsequently processed in KNIME as soon as the complete analysis finished. All features were initially aligned based on their mass and retention time. The resulting consensus list contained about 30000 unique analytes. Subsequent data filtration steps (e.g. fold change, intensity tresholds) reduced the dataset to about 2200 significant features. Multivariate statistical tools (e.g. PCA, hierarchical clustering) were applied for data evaluation and visualization. Furthermore, inclusion lists were automatically generated for subsequent MS-MS experiments for analyte identification.