AUTOMATIC DETECTION OF CONCENTRATION TRENDS OF ORGANIC POLLUTANTS IN WASTEWATER USING COMPUTATIONAL APPROACHES AND CHEMOMETRIC TOOLS ON DATA ACQUIRED BY LC-HRMS

ALYGIZAKIS N.A.[1], GAGO FERRERO P.[1], LOOS M.[2,3], SINGER H. [2], HOLLENDER J.[2,3] and THOMAIDIS N.S.[1]

[1] National and Kapodistrian University of Athens, Department of Chemistry, Laboratory of Analytical Chemistry, Panepistimiopolis 157 71, Athens, Greece, [2] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland, [3] Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland
E-mail: ntho@chem.uoa.gr

Contaminants of emerging concern enter daily into the sewage system in large quantities. Concentration levels of the different compounds are not constant, and may follow different trends and patterns, affected by several factors (like usage trends, or pollution spills). The investigation of the concentration pattern of both target and non-target substances in wastewater is an important issue, as it could be used as an early-warning system on pollution loads, it may provide a further insight on the behaviour of organic contaminants in the environment, and also information about the community use of chemicals. The objective of the presented study is to develop an automated workflow for the detection of differences and trends in intensities for all the detected (target and non-target) substances among different sampling sets (e.g. time periods, influent/effluent) through the use of a workflow based on algorithms from the packages XCMS and CAMERA and use of the web based platform MetATT.

Temporal sequences of samples were collected from the wastewater treatment plant of Athens. Analysis was carried out by liquid chromatography - high resolution tandem mass spectrometry (LC-Q-TOFMS). First, the acquired data is converted to mzXML using ProteoWizard [1] and stored in two subfolders (e.g. influents and effluents) in the R working folder. Sample feature detection is performed by the centWave algorithm with optimized parameters for QTOF MS data. Features representing the same analyte across samples are placed into groups, peak alignment follows, missing features are filled with a low intensity value [2,3], and then, features are clustered according to peak shape correlation coefficients and retention times. For this purpose, *xcmsSet* object is converted to CAMERA object. Finally, isotopic peaks and adducts are annotated to the same chemical component and its monoisotopic peak [4]. The peak list is exported in a .csv file to the web-based application MetATT (http://metatt.metabolomics.ca/MetATT/). This platform uses Multivariate Empirical Bayes Approach [5] for providing prioritized peaks (using Hotelling T2 coefficient) according to the differences of integrated intensities among the two studied groups. Moreover, it automatically produces plots to evaluate the trend of each substance. After that, the most relevant peaks can be tentatively explored using non-target identification approaches. The presented approach enables the recording of concentration trends for a large number of compounds in a given set of samples. Moreover, this workflow can be used to detect events of direct disposal of some specific substances into the sewage system, constituting an appropriate source of information for WWTP authorities.

Keywords: trends analysis, wastewater, R-project, emerging contaminants, computational tools

ACKNOWLEDGMENTS

This research has been co-financed by the European Union and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference

Framework (NSRF) – ARISTEIA 624 (TREMEPOL project). N.A. Alygizakis would like also to thank John S. Latsis Foundation for the financial support.

REFERENCES

1. Kessner D., Chambers M., Burke R., Agus D., Mallick P. (2008), ProteoWizard: open source software for rapid proteomics tools development., Bioinformatics, 24(21), pp. 2534-2536, DOI: 10.1093/bioinformatics/btn323
2. Smith C., Want E., O'Maille G., Abagyan R. and Siuzdak G. (2006), XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification., Analytical Chemistry, 78(3), pp. 779–787, DOI: 10.1021/ac051437y
3. Tautenhahn R., Boettcher C. and Neumann S. (2008), Highly sensitive feature detection for high resolution LC/MS., BMC Bioinformatics, 9(504), pp. 504, DOI: 10.1186/1471-2105-9-504
4. Kuhl C., Tautenhahn R., Böttcher C., Larson T., and Neumann S. (2012), CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets., Analytical Chemistry, 84(1), pp. 283-289, DOI: 10.1021/ac202450g
5. Xia J., Sinelnikov I. and Wishart D. (2011) , MetATT: a web-based metabolomics tool for analyzing time-series and two-factor datasets., Bioinformatics, 27(17), pp. 2455-2456, DOI: 10.1093/bioinformatics/btr392