Autonontarget: An R Package to Perform Automatic Non-Target Screening

Aalizadeh R., Thomaidis N. S [*]

Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Panepistimiopolis Zografou, 15771 Athens, Greece

*E-mail: ntho@chem.uoa.gr

Abstract

Recent advances on Liquid Chromatography-High Resolution Mass Spectrometry (LC-HRMS) has revolutionized identification of new compounds, having various polarity, over various scientific fields especially in the environmental science. The continuous growing of LC-HRMS applications yet increased the "peak inventories". This is achieved within three main workflows of "Target", "Suspect" and "Non-target" screening. Although targeted analysis is the best way to confirm the identification of a compound, it is sometime not practical due to limited access to reference standards. The vast majority of the peaks detected in the samples generally remain unidentified and supportive information such as retention time prediction, MS/MS (experimental and estimated one) along with ionization behavior would help increase the identification confidence. As "peak inventories" expands and number of regulatory databases grows, retrieving possible candidates and screening them often become a time-consuming task and requires large amount of efforts. Thus, an automatic approach could be of great need to screen known-unknown compounds in the samples. The aim of this study is to propose a workflow (Automatic Non-target Screening (AutoNonTarget)) to screen a peaks-list, created by LC-HRMS instrument, from the environmental samples such as influent/effluent wastewater (IWW and EWW) or sewage sludge samples.

The proposed workflow starts with an optimized peak picking algorithm, using XCMS and enviPick, for a set of samples (16 IWW, EWW and sewage sludge samples) in which the MS and MS/MS information were recorded in data independent/dependent acquisition mode. After deriving a peaks-list, the peaks (*m/z*) originated from analytical procedural blank is subtracted from the sample using set an advanced chemometric method. Then, each remaining annotated peak (mainly, but not exclusively the molecular ions, adducts, double charged ions etc. detected by CAMERA and non-target R packages) or mass of interest (an ion that has significant fold changes from one sample to another, trend, or high loading weights) is searched within publicly available database (such as PubChem, REACH database, FoodB or user defined database etc.) and corresponding candidates (within a certain mass accuracy (mD/ppm) provided by user) are being retrieved afterwards. Next, the theoretical isotopic pattern is calculated for each candidate by "enviPat" and then compared with extracted experimental isotopic pattern [1]. Then, the experimental and predicted retention time indices of each candidate is derived to filter out false positives and support identification subsequently. AutoNonTarget includes additional steps to incorporate the MS/MS information during identification such as use of MetFrag [2] (for interpreting the MS/MS fragments and derive score) and CFM-ID (not only to annotate the MS/MS fragments for a list of candidates, but also to predict MS/MS spectrum at various collision energies (CE)) [3]. Finally, the public and open access mass spectral library is used by AutoNonTarget to verify the identification of the compounds. A new method is also used to calculate MS/MS similarity score taking into account effect of different collision energies between the MS^2 spectra of candidates and reference standards (found in the mass spectrum libraries). AutoNonTarget provides a final table including all these information with derived level of identification confidence [4] for each *m/z* to facilitate the identification task in given samples by suspect/non-target screening. The workflow is evaluated by set of 100 common emerging contaminants (prepared in a solution of 50:50 (H_2O:MeOH) by their reference standards), treated as unknown, and applied externally for screening of detected *m/zs* in the Norman SusDat (https://www.norman-network.com/?q=node/236) and PubChem database (https://pubchem.ncbi.nlm.nih.gov/).

Analyses of wastewater samples were carried out by UHPLC-QToF-MS. More details about the analytical method used can be found in [5]. Among the compounds detected by AutoNonTarget, several pharmaceuticals, personal care products, disinfectants and surfactants, like 2-[2-(3-aminopropoxy)ethoxy]ethanol and tetraethylene glycol were identified. "AutoNonTarget" greatly facilitates a higher confidence, rapid screening of samples for suspects, providing an overview of tentative identifications and likely false positive matches for subsequent follow-up.

References
[1] M. Loos et al., Analytical Chemistry, (2015) 87, 5738-5744.
[2] C. Ruttkies et al., Journal of Cheminformatics, (2016) 8, 3.
[3] F. Allen et al., Nucleic Acids Research, 42 (2014) W94-W99.
[4] E.L. Schymanski et al., Environmental science & technology, (2014) 48 , 2097-2098.
[5] P. Gago-Ferrero et al., Environmental Science & Technology, (2015) 49(20), 12333-12341.