Open-source workflow for smart biotransformation product elucidation using LC-HRMS data

Stravs M.A.[1,2]*, And Hollender J.[1,2]

[1]Eawag, Department of Environmental Chemistry, Überlandstr. 133, 8600 Dübendorf, Switzerland

[2]ETH Zürich, Institute of Biogeochemistry and Pollutant Dynamics, Univerisätstrasse 16, 8092 Zürich, Switzerland

*corresponding author:

e-mail: stravsmi@eawag.ch

Abstract *RMassScreening* (https://github.com/meowcat/RMassScreening) is presented, a workflow for the elucidation of micropollutant transformation products based on open-source tools. The workflow combines script-based processing with interactive data exploration. The workflow was applied to the elucidation of biotransformation reactions in phytoplankton, finding 14 transformation products for 9 micropollutants, and used for the exploration of biodiversity effects on total biotransformation potential.

Keywords: HRMS, data processing, computational mass spectrometry, transformation products

1. Introduction

In recent years, knowledge about occurrence and fate of micropollutants in the environment has progressed from both the expansion of targeted screening approaches with large numbers of new substances, and suspect and non-target screening for unknown compounds. Using high-resolution tandem mass spectrometry (HRMS/MS), comprehensive information about organic micropollutants in samples can be acquired with a single or few complementary measurements. However, only a minor subset of the rich data in such measurements can currently be assigned to chemical entities. It is believed that a large number of unidentified chemical signals arises from biotic or abiotic transformation of known micropollutants. Therefore, the understanding of biotransformation processes is crucial to gain a more comprehensive picture of micropollutant behavior in the environment.

Transformation products (TPs) of micropollutants can be identified through prediction of possible transformation reactions from parent compounds, or found via screening for time profile behavior in biotransformation experiments. Commercial software is available for the purpose of transformation product elucidation (Bletsou *et al.*, 2015), which is often vendor-specific with limited flexibility and provides limited possibilities for interaction with other software. Alternatively, freely available and powerful open-source tools can be used in combination for the same purpose. Complete open-source HRMS/MS processing workflows exist, particularly for metabolomics data analysis (Edmands *et al.*, 2017) or environmental non-target screening with spill detection (https://github.com/blosloos/enviMass). However, for the purpose of TP elucidation, no open-source integrated workflows exist to date.

Herein, RMassScreening, a workflow for TP elucidation is presented (https://github.com/ meowcat/RMassScreening). It integrates processing using a combination of open-source tools, and visual interactive analysis of the results. Two exemplar applications on biotransformation experiments with phytoplankton are shown.

2. Results

2.1. Workflow

The workflow *RMassScreening* is comprised of a script-based processing stage and an interactive graphical data analysis and exploration stage (Figure 1, top). For data processing, the user specifies:

- an input set of raw, centroided LC-HRMS data files in mzXML format
- a set of parent compounds, specified by molecular formula or by exact mass, as a CSV table
- a set of candidate reactions, specified by molecular formula modifications or exact mass shifts, to be applied iteratively on the parent compounds
- an assignment of raw files to samples and timepoints, in CSV format
- an assignment of samples to experimental groups with different properties in CSV format.

In the script-based processing stage, peak picking on raw files is performed using enviPick (https://github.com/blosloos/enviPick) and peaks are aligned across samples to form profiles using either enviMass (http://github.com/blosloos/enviMass) or XCMS (Smith *et al.*, 2006) backends. To associate isotopes, adducts and in-source fragment peaks to components, an optimized version of RAMClustR (Broeckling *et al.*, 2014) was specifically developed which is able to handle a large number of components using hierarchical clustering on sparse matrices (https://github.com/meowcat/RAMClustR, https://github.com/meowcat/fastliclust).

The workflow provides functions to build a suspect list from parent compounds and candidate reactions, which can be applied consecutively multiple times, generating large suspect lists. Alternatively, suspect lists can be imported from any other tabular output, e.g. from pathway prediction systems. The profiles can subsequently be screened for potential TP candidates, and the results compiled into time series by experimental groups and conditions.

Suspect screening results are finally loaded in an interactive viewer application, which allows to filter and sort results according to custom criteria based on ratios between different experimental groups, conditions and/or timepoints (Figure 1, bottom left). For each entry, the time profiles can be directly reviewed (Figure 1, bottom right). Instead of suspect screening results, other result lists obtained through e.g. non-target priorization methods can also be loaded into the viewer.

2.2. Applications

The workflow was applied for the identification of biotransformation products in phytoplankton organisms. In batch experiments with three species and 24 micropollutants, 14 TPs were found for 9 parent compounds, whereas no TPs were found for the other parents (Stravs *et al.*, 2017). The observed reactions included hydrolysis reactions, redox reactions and conjugation reactions. In addition to common biotransformation reactions, some reactions apparently performed through promiscuous enzymes were observed, such as sulfamethoxazole pterin conjugation and mefenamic acid glutamate conjugation.

In a different study, 27 combinations of phytoplankton species from five different phytoplankton functional groups were assembled with differing species richness and phytoplankton functional group richness, each in duplicate. The assemblages were exposed to a mixture of 37 compounds. Through a broad-range suspect screening with the developed tool, the number of formed TPs was determined for each experiment. The number of total TPs was correlated with both species richness and functional group richness, indicating a potential role of natural communities biodiversity in biotransformation of polar organic micropollutants.

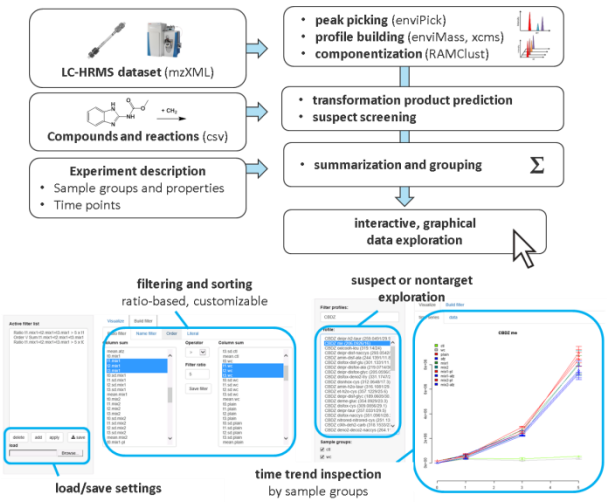

Figure 1. Top: Schematic processing workflow. Bottom: Exemplary commented screenshots: interactive filter generation (left), data analysis and exploration (right)

References

Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. TrAC Trends Anal. Chem. 66, 32–44. doi:10.1016/j.trac.2014.11.009

Broeckling, C.D., Afsar, F.A., Neumann, S., Ben-Hur, A., Prenni, J.E., 2014. RAMClust: A novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal. Chem. 86, 6812–7. doi:10.1021/ac501530d

Edmands, W.M.B., Petrick, L., Barupal, D.K., Scalbert, A., Wilson, M.J., Wickliffe, J.K., Rappaport, S.M., 2017. compMS2Miner: An Automatable Metabolite Identification, Visualization, and Data-Sharing R Package for High-Resolution LC–MS Data Sets. Anal. Chem. 89, 3919–3928. doi:10.1021/acs.analchem.6b02394

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. 78, 779–87. doi:10.1021/ac051437y

Stravs, M.A., Pomati, F., Hollender, J., 2017. Exploring micropollutant biotransformation in three freshwater phytoplankton species. Environ. Sci. Process. Impacts 19, 822–832. doi:10.1039/C7EM00100B