Identifying Complex Mixtures in the Environment with Cheminformatics and Non-targeted High Resolution Mass Spectrometry

Emma L. Schymanski[1,2]*, Antony J. Williams[3], Juliane Hollender[2]

[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg.
[2]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland.
[3]National Center for Computational Toxicology, US EPA, Research Triangle Park, Durham, NC, USA.

* emma.schymanski@eawag.ch

Non-targeted high resolution tandem mass spectrometry (HR-MS/MS) techniques are ideal for exploring complex environmental samples beyond target and suspect screening for "knowns". Non-target HR-MS/MS, combined with advanced cheminformatics and data processing, offers huge potential for exploring complex mixtures, yet faces many challenges. Peak inventories of several non-target studies reveal that only a few percent of detected masses are "known" compounds, while several thousand peaks remain unidentified. Mass-based homologue screening shows that many of these peaks are interconnected through a web of related, discrete masses. Homologue screening in wastewater studies confirms this for several surfactant series. A better understanding of all "unknowns" in samples, including homologues, will be essential to unravel the role that complex mixtures play in environmental observations.

Improved chemical identification is hampered by the difficulty in exchanging information on substances, yet incorporating chemical knowledge and fragment information is vital to unravelling the identities of these "unknowns" and assessing the environmental and toxicological relevance of mixtures. The collaborative efforts between the US EPA's CompTox Chemistry Dashboard (https://comptox.epa.gov/) and the NORMAN Network's Suspect Exchange (http://www.norman-network.com/?q=node/236) and MassBank (www.massbank.eu) initiatives showcase how improved information exchange can help address this challenge. Connecting these initiatives into non-target workflows and *in silico* structure elucidation approaches such as MetFrag (https://msbi.ipb-halle.de/MetFragBeta/) allows rapid tentative identification of potential candidates amongst thousands of non-target masses. These efforts will be showcased on European studies, showing how non-target HR-MS helps turn the "unknowns" into "knowns" for improved assessment of complex mixtures in the environment. *Note: This abstract does not reflect US EPA policy*.