Environmental Cheminformatics to Identify Unknown Chemicals and their Effects

ABSTRACT:
The environment and the chemicals to which we are exposed is incredibly complex, with over 125 million chemicals registered in the largest chemical registry and over 70,000 in household use alone. Detectable molecules in complex samples can now be captured using high resolution mass spectrometry (HRMS), which provides a “snapshot” of all chemicals present in a sample and allows for retrospective data analysis through digital archiving. However, scientists cannot yet identify the vast majority of the tens of thousands of features in each sample, leading to critical bottlenecks in identification and data interpretation. For instance, recent studies indicate a strong connection between the gut microbiome and Parkinson’s disease, yet over 60 % of significant metabolites in microbiome experiments are unknown. Unknown identification remains extremely time consuming and, in many cases, a matter of luck. Prioritizing efforts to find significant metabolites or potentially toxic substances responsible for observed effects is the key, which involves reconciling highly complex samples with expert knowledge and careful validation. This talk will cover European, US and worldwide community initiatives to help connect knowledge on chemistry and toxicity with environmental observations - from compound databases to spectral libraries and retrospective screening. It will touch on the challenges of standardized structure representations, data curation, deposition and communication between resources. Finally, it will show how interdisciplinary efforts and data sharing can facilitate research in metabolomics, exposomics and beyond.