Community Resources Connecting Chemistry and Toxicity Knowledge to Environmental Observations

Emma Schymanski[1], Antony Williams[2]

Luxembourg Centre for Systems Biomedicine, Luxembourg, Luxembourg[1], National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Durham,USA[2]

Exposure to chemicals may be a causative or contributing factor to the progression of diseases. On the other hand, chemicals such as pharmaceuticals are also used to prevent, treat or alleviate the symptoms of diseases. Considering a far broader range of small molecules than are currently captured in many biological networks is a huge challenge for disease maps. The number of chemicals for consideration is daunting. The largest chemical databases presently available contain ~100 million chemicals (of which many were never produced in significant amounts), while Europe produces or imports >143,000 substances above 1 tonne/yr, and an estimated 70,000 chemicals are used in households. The latest version of the Human Metabolome Database (HMDB) now contains 114,100 metabolites, yet many of these are predicted structures. Smaller resources such as the US EPA's environmentally focused CompTox Dashboard contain >760,000 chemicals, but even this collection includes chemicals such as isotopically labeled compounds and "UVCB chemicals" - unknown or variable composition, complex reaction products or biological materials, which can represent 10s to 100s of individual chemical components. However, all of these chemical databases are inherently incomplete. In high resolution mass spectrometry (HR-MS) measurements, used to measure chemicals in metabolomics, exposomics, foodomics, forensics and personalized medicine, we are confronted with tens of thousands of features, of which only a few percent can be annotated as "known" and confirmed as metabolites or chemicals of interest using all available chemical databases. How can we reconcile our chemical knowledge with our sample observations? This talk will cover European, US and worldwide community initiatives to help connect knowledge on chemistry and toxicity with environmental observations, i.e. helping researchers to find small molecules in big data in smarter ways - 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, aiming to stimulate discussions on possibilities for integrating these approaches into disease maps.
Note: this abstract does not necessarily represent U.S. EPA policy.