ACToR Chemical Structure processing using Open Source ChemInformatics Libraries

Jayaram Kancherla[1], Kamel Mansouri[1], Hao Truong[1], Ann M. Richard[1], Richard S. Judson[1]

[1] NCCT/ORD/USEPA, RTP, NC 27711

ACToR (Aggregated Computational Toxicology Resource) is a centralized database repository developed by the National Center for Computational Toxicology (NCCT) at the U.S. Environmental Protection Agency (EPA). Free and open source tools were used to compile toxicity data from over 1,950 public sources. ACToR contains chemical structure information and toxicological data for over 558,000 unique chemicals. The database primarily includes data from NCCT research programs, in vivo toxicity data from ToxRef, human exposure data from ExpoCast, high-throughput screening data from ToxCast and high quality chemical structure information from the EPA DSSTox program. The DSSTox database is a chemical structure inventory for the NCCT programs and currently has about 16,000 unique structures. Included are also data from PubChem, ChemSpider, USDA, FDA, NIH and several other public data sources. ACToR has been a resource to various international and national research groups. Most of our recent efforts on ACToR are focused on improving the structural identifiers and Physico-Chemical properties of the chemicals in the database. Organizing this huge collection of data and improving the chemical structure quality of the database has posed some major challenges. Workflows have been developed to process structures, calculate chemical properties and identify relationships between CAS numbers. The Structure processing workflow integrates web services (PubChem and NIH NCI Cactus) to download structure information and uses RDKit to calculate structural properties. Workflows that process all structures available in the database to make them QSAR-ready and find different structural relationships are also in progress. All ACToR data is publicly downloadable and our current efforts are focused on making a reliable source of chemical and toxicity information. *This abstract does not necessarily reflect U.S. EPA policy*