(IIB-1) DSSTox Project: Supporting Improved Toxico-Chemoinformatics Capabilities

Authors: Ann Richard[1], Maritja Wolf[2], Jamie Burch[3], ClarLynda Williams[3], Kun Wang[4] and Alexander Tropsha[4]
[1]National Center for Computational Toxicology, U.S. EPA, RTP, NC, [2]Lockheed Martin, Contractor to the EPA, [3]EPA NCSU Graduate Student COOP – Bioinformatics Graduate Program, [4]EPA/NCCT-UNC Star Grant Bioinformatics Center

Major trends affecting public toxicity information resources have the potential to significantly alter the future of predictive toxicology. Chemical toxicity screening is undergoing shifts towards greater use of more fundamental information on gene/protein expression patterns and bioactivity and bioassay profiles, the latter generated with high-throughput screening technologies. Curated, systematically organized, and web-accessible toxicity and biological activity data in association with chemical structures, enabling the integration of diverse data information domains, will fuel the next frontier of advancement for QSAR (quantitative structure-activity relationship) and data mining technologies. The DSSTox project is supporting progress towards these goals on many fronts, promoting the use of formalized and structure-annotated toxicity data models, helping to interface these efforts with QSAR modelers, linking data from diverse sources, and creating a large, quality reviewed, central chemical structure information resource linked to various toxicity data sources. As the DSSTox project and corresponding Master File have grown and branched into new areas, we have implemented a variety of measures to ensure consistent data quality across all DSSTox Structure Data (SD Format) Files, to automate data entry and review, to facilitate construction of new DSSTox SDF files, to enable comparison of chemical content across SDF files, to serve as a central data indexing resource for chemical structures information across EPA programs, and to interface with the large NIH PubChem resource. In addition, new DSSTox projects to support the next generation QSAR approaches are exploring the use of deeper, more elaborated biological activity descriptions as a means for projecting chemical property space more appropriately into the biological domain. These various efforts will be illustrated by collaborations among the Carcinogenic Potency Project, NCI, NIH Molecular Libraries Initiative and Chemical Genomics Center, PubChem, Leadscope ToxML, various EPA programs, and the National Toxicology Program.