Abstract for presentation at Health Canada's "New Substances Assessment and Control Bureau - All Staff Science Day", Ottawa, Canada, 21 November 2007

Recent Developments in Toxico-Cheminformatics: A New Frontier for Predictive Toxicology

Ann Richard
National Center for Computational Toxicology, US EPA, RTP, NC 27711; richard.ann@epa.gov

Efforts to improve public access to chemical toxicity information resources, coupled with new high-throughput screening (HTS) data and efforts to systematize legacy toxicity studies, have the potential to significantly improve predictive capabilities in toxicology. Important recent developments include: 1) large and growing public resources that link chemical structures to biological activity and toxicity data in searchable format, and that offer more nuanced and varied representations of activity; 2) standardized relational data models that capture relevant details of chemical treatment and effects of published in vivo experiments; and 3) the generation of large amounts of new data from public efforts that are employing HTS technologies to probe a wide range of bioactivity and cellular processes across large swaths of chemical space. Chemical structure is effectively linking data across diverse study domains (e.g., 'omics', HTS, traditional toxicity studies), toxicity domains (carcinogenicity, developmental toxicity, neurotoxicity, immunotoxicity, etc) and database sources (EPA, FDA, NCI, PubChem, GEO, ArrayExpress, etc.). Most recently, EPA's DSSTox project has published several new EPA chemical data inventories (IRIS, HPV, ToxCast) and added an on-line capability for structure (substructure or similarity)-searching through all or parts of the published DSSTox data files. These efforts are, for the first time in many cases, opening up a structure-paved two-way highway between previously inaccessible or isolated public chemical data repositories and large public resources, such as PubChem. In addition, public initiatives (such as ToxML) are developing systematized data models of toxicity study areas, and introducing standardized templates, controlled vocabularies, hierarchical organization, and powerful relational searching capability across newly captured data. Cheminformatics and data models, in turn, are providing the underpinning for the large public HTS efforts of the NIH Molecular Libraries Initiative, as well as new toxicity-targeted HTS programs within the EPA and the NIEHS National Toxicology Program. These initiatives are turning the structure-activity paradigm on its head, using chemicals to probe biological space and generating "biological profiles" of chemicals that, along with chemical structure considerations, offer the promise of providing richer, and more relevant and predictive associations to in vivo responses. *This work was reviewed by EPA and approved for publication, but does not necessarily reflect EPA policy.*