

233rd ACS National Meeting, Chicago, IL, March 25-29, 2007

Applying data mining approaches to further understanding chemical effects on biological systems

Chihae Yang (1) and Ann M. Richard (2)

(1) Leadscope, inc. 1393 Dublin Road, Columbus, OH 43215

(2) National Center for Computational Toxicology, US EPA, Research Triangle Park, NC 27711

Data mining methods require the technological framework of a relational database based on a rigorous data model, flexible searching and retrieval functions, and data analysis and visualization tools. A data model, consisting of a schema (hierarchy) and controlled vocabulary, provides the foundation for meaningful data mining, enabling mechanistic hypotheses to be generated and validated. Advances in the field of computational toxicology are being driven by expanding capabilities for mining the domains of biology and chemistry simultaneously. To break away from the current paradigm of analog searching solely based chemical similarity, this paper presents informatics methods to finding chemical structures with biologically similar functions. A chemical stressor with particular biological attributes will seed the biology domain. The resulting biological profile will then be projected onto the chemical structure domain to broaden the concept of “analogs” and to assist in the understanding of hazard potential through iterative exploration of both chemical and biological analog space. The National Toxicology Program recently conducted high throughput screening of over 1400 chemicals in a series of cell-viability assays and made the data available through PubChem. This dataset will be used to illustrate various data mining techniques to biologically profile the chemical space.

This abstract does not necessarily reflect EPA policy.