

Carcinogenicity and Mutagenicity Data: New Initiatives to Improve Access & Utility for Modeling

Ann Richard, National Center for Computational Toxicology, US EPA, Research Triangle Park, NC 27711, Richard.ann@epa.gov

Current models for prediction of chemical carcinogenicity and mutagenicity rely upon a relatively small number of publicly available data resources, where the data being modeled are highly summarized and aggregated representations of the actual experimental results. A number of new initiatives are underway to improve access to existing public carcinogenicity and mutagenicity data for use in modeling, as well as to encourage new approaches to the use of data in modeling. Rodent bioassay results from the NIEHS National Toxicology Program (NTP) and the Berkeley Carcinogenic Potency Database (CPDB) have provided the largest public data resources for building carcinogenicity prediction models to date. However, relatively few and limited representations of these data have actually informed existing models. Initiatives, such as EPA's DSSTox Database Network, offer elaborated and quality reviewed representations of the CPDB, and expanded data linkages and coverage of chemical space for carcinogenicity and mutagenicity. In particular, the latest published DSSTox CPDBAS structure-data file includes a number of species-specific and summary activity fields, including a species-specific normalized score for carcinogenic potency (TD50), and various weighted summary activities. These data are being incorporated into PubChem to provide broad "read-across" in bioassay and structure space. Additionally, collaborations are underway to publish more complete summary versions of NTP carcinogenicity and genetic toxicity data. Finally, as part of EPA's ToxCast program, detailed data from rodent bioassay studies are being collected in standardized representations for hundreds of regulated pesticides, and various representations of these data will be made publicly available. These data are also being used to anchor "next-generation" predictive technologies, which are based on high-throughput screening data used in conjunction with chemical structure information. *This abstract does not reflect EPA policy.*