DSSTox Project: Supporting Improved Toxico-Chemoinformatics Capabilities

Ann Richard[1], Maritja Wolf[2], Jamie Burch[3], ClarLynda Williams[3], Kun Wang[4] and Alexander Tropsha[4]

[1]U.S EPA, ORD, Computational Toxicology Program.; [2]Lockheed Martin, Contractor to the EPA, [3]EPA-NCSU Student COOP, [4]EPA/NCCT-UNC Star Grant Bioinformatics Center

research&development

Project ID: IIB-1

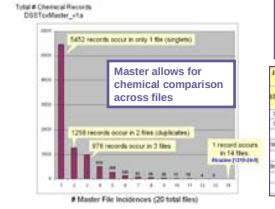
Science Question

Chemical structure is a universal information metric that cuts across traditional domains of toxicological investigation, and empowers the concept of "chemical analogy" in support of biological extrapolation and inferences. Curated, systematically organized, and web-accessible toxicity and biological activity data in association with chemical structures will enable the integration of diverse data information domains, and will fuel the next frontier of SAR (structure-activity relationship) and data mining technologies.

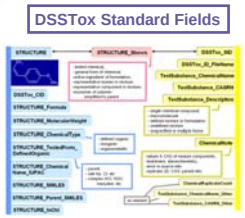
What tools and capabilities must be developed to reach these objectives?

Research Goals

- Develop and promote the use of chemical information and file format standards that can accommodate diverse toxicity data content;
- Coordinate with internal and outside efforts to standardize chemical and toxicity-related information, and encourage data integration and data mining capabilities across public toxicity databases;
- Collaborate with toxicity source experts to define appropriate levels and description of summary toxicity measures that can enrich options for data mining, hypothesis generation, and SAR modeling;
- Create a high quality central structure-index for diverse EPA databases and chemical lists that are important to EPA's regulatory mission, constructed to support data integration efforts both within and outside EPA.



DSSTox Standard Fields

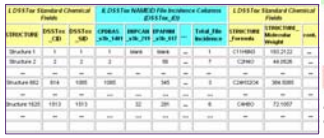


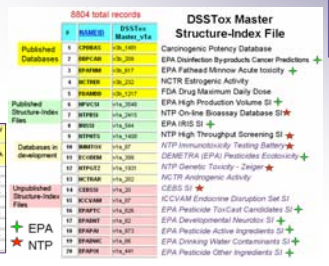
Published Structure-Data Files

- Structure-Index Files (standard chem fields only)
- Structure-Index Locator Files (SI + links to data)
- Database Files (SI + toxicity data)

DSSTox Master Structure-Index File

- Central consolidation of all DSSTox chemical records
- Chemical information QA procedures
- File Incidence Columns allow file content comparison
- Central database for EPA public structure browser
- Downloadable as SDF from DSSTox website
- Enable chemoinformatic analysis of component DBs





http://www.epa.gov/ncct/dsstox/

DSSTox Website

- Publish Structure-Data Files
- Publish Standards, QA procedures
- Central information resource
- Documentation, links to Source info
- Coordinate public efforts

DSSTox QA Procedures

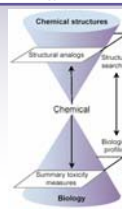
- Resolve Chemical names & CAS
- Multiple public sources – consensus
- Use primary sources if unresolved errors
- Structure accuracy & consistent representation
- automated scripts for error checking

Coordinating Public Efforts

DSSTox

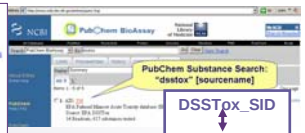
Toxicity Data Models to Build Data Mining Capabilities

- ToxML Coordination
- ILSI Developmental Toxicity Database Workgroup
- EPA Developmental Neurotox Database
- EPA ToxCast Pesticides Database
- Immunotoxicology Database



PubChem Collaboration

- Publish all DSSTox Structure-Data Files
- Update DSSTox standard fields (CID,SID)
- DSSTox "depositor" associated with DSSTox QA
- Structure-searching through all DSSTox files
- Linkage to >5 million PubChem bioassay records

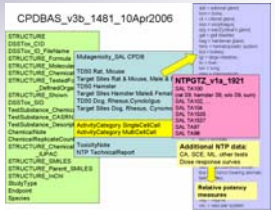


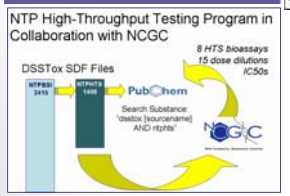
NTPHTS/NCGC/PubChem

- NTPHTS SI file → PubChem → NCGC
- NCGC HTS Bioassay data → PubChem NTPHTS
- PubChem NTPHTS → DSSTox → NTP

Activity Representations for SAR modeling & Data Mining

- Aggregated activity levels
- Summary activities, dose response, potency
- Integration of bioassay profiling data



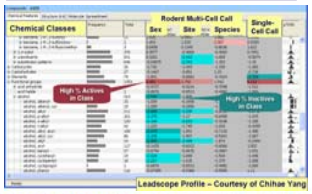


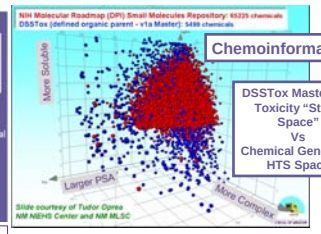
EPA NCCT ToxCast Collaboration

- Compile pesticide actives structure files
- Assess overlaps with other structure files (HPV, NTP, IRIS, etc)
- Provide central structure-indexing ToxCast database and link-out to public toxicity reference data
- Eventually publish companion summary toxicity data files

Bioassay Profiling

- Collaboration with NCI
  - generate PASS bioactivity predictions for entire DSSTox Master File
- Collaboration with UNC (K Wang, A Tropsha)
  - utilize different types of activity measures with chemical functional hierarchies (Leadscope) to explore new SAR approaches to prediction





Chemoinformatics

DSSTox Master File Toxicity "Study Space" Vs Chemical Genomics HTS Space

Progress & Results

- Major DSSTox website update live on 12 June 2006
- Publish Master file with major QA review
- Coordinate with PubChem to revise standard fields, publish DSSTox files, facilitate interactions with NTP
- Expand collaborations with NTP, NTPHTS and NCGC
- Expand collaborations within EPA (ToxCast, Chemoinformatics CoP, Green Chemistry, HPV)
- Initiated project to create EPA-wide chemical structure browser based on DSSTox Master file and component data files
- Expand collaborations outside of EPA (NCI, UNC)
- Increase involvement in outside data model efforts (ToxML, ILSI)

Impact and Outcomes

- DSSTox is becoming synonymous with accurate, quality chemical structures and annotation associated with toxicity databases.
- Published DSSTox Structure-Data Files are encouraging greater attention to chemical information quality and structure-annotation
- DSSTox collaborations with outside parties (PubChem, ToxML, NCGC, etc) are creating linkages of EPA data to outside data resources

Future Directions

- DSSTox standards and Master File provide foundation for building a common, central structure-index and on-line browser for integrating EPA chemical data
- Expand DSSTox to data domains of toxicology not previously reached by data models (immunotox, develop. tox, develop. neurotox, etc)
- Expand involvement in EPA ToxCast effort, using DSSTox as major integrator to outside reference toxicity data
- Explore new ways of summarizing and representing biological activity to expand SAR and data mining capabilities

Acknowledgements

Ray Tice, Cynthia Smith, Beth Bowden, Brad Collins – NIEHS/NTP
David Dix, Keith Houck, Matt Martin – ToxCast
Chihae Yang – Leadscope
Steve Bryant, Jane Tseng, Yanli Wang – PubChem Project
Jennifer Fostel – NIEHS CEBS project
Lois Gold – Carcinogenic Potency Project
Marc Nicklaus – National Cancer Institute
Tudor Oprea – Univ. of New Mexico, School of Medicine
Chris Austin, Ajit Jadhav (NIH/NCGC); Doug Livingston (DPI-SMR)

Long-term goal II

This poster does not necessarily reflect EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.