The EPA Comptox Chemistry Dashboard

NCCT: Antony Williams, Chris Grulke, Jennifer Smith, Kamel Mansouri, Andrew McEachran, Ann Richard and Jeff Edwards

NERL: Jon Sobus, Kathie Dionisio and Katherine Phillips

NRMRL: Todd Martin

United States Environmental Protection Agency

Science Context

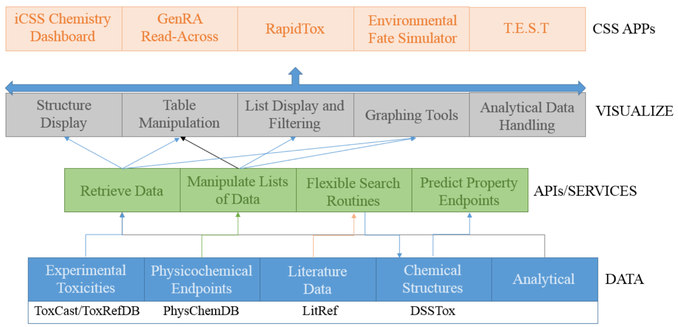
A consolidated web platform is necessary for researchers to access chemical information look-up, models and model predictions and linkages to Agency and public resources. This will provide access to: curated chemical structures, computed and measured physchem properties, exposure and usage patterns, data analysis workflows, reaction chemistry, and Quantitative Structure-Activity Relationship (QSAR)-based physiologically based pharmacokinetic (PBPK) and toxicity predictions. The web platform will be extended to include additional capabilities to support "read-across" approaches. Such resources will be made available for broadest possible usage by cheminformaticians, modelers, non-chemists and risk assessors.

The views expressed in this poster are those of the author[s] and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Approach

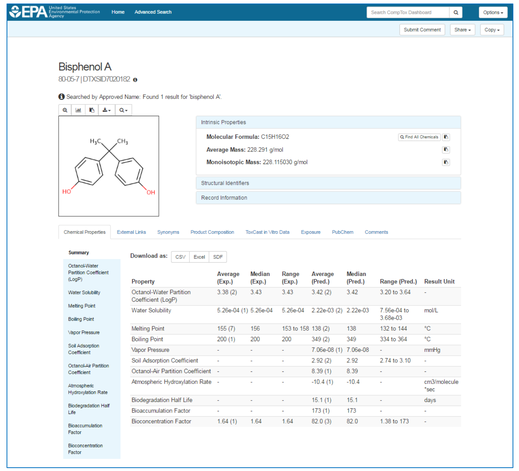
1) Develop underlying componentized architecture utilizing a "build once, use many times" approach

A Block Diagram of the Cheminformatics Architecture

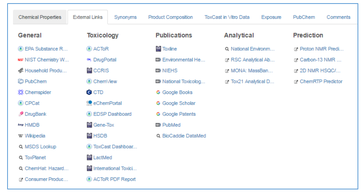


2) Deliver web application access to DSSTox database (ca. 720,000 structures). Beta version delivered April 2016 at ACS Spring.
3) Integrate other data via web services (CPCat, Toxcast, ExpoCast data) for Version 1 release, August 2016 at ACS Fall Meeting.
4) Utilize Chemistry Dashboard as underpinning architecture to develop RapidTox prototype by August 2016
5) Demonstrate value of dashboard to support "Non-Targeted Analysis" to examine the Exposome (*publication submitted*)

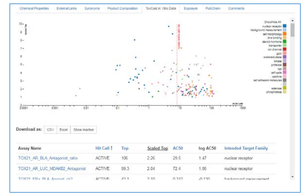
Preliminary Results



The chemical page for Bisphenol A. Physical Properties Tab listing experimental and predicted properties.

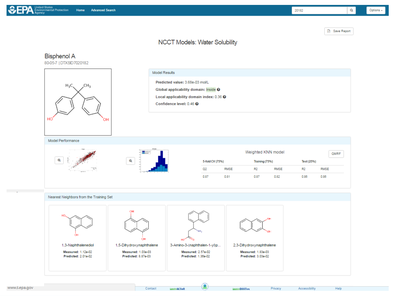


External Links to EPA, interagency and public websites



Toxcast Screening Data

Functional Use and Composition Data (from CPCat)



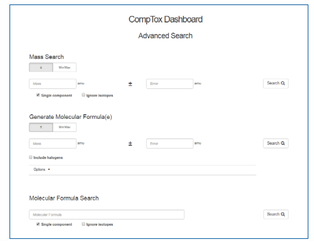
Detailed Model Reports:

- Predicted data including logP, water solubility and Bioconcentration Factor
- Algorithms developed using data from PHYSPROP dataset
- Data curated using combined manual process & automated workflow[1]
- Model Reports show global and local applicability domain and nearest neighbors in the training set
- "QSAR Modeling Report Format" (QMRF) reports

Non-Targeted Screening Support Through Mass and Formula-based Searching

- Support identification of chemicals in different environmental media (e.g. dust, water) utilizing mass spectrometry based non-targeted screening
- Both mass- and formula-based searches are supported, single and batch
- Utilize rank ordering of hit lists by the number of data sources, functional use and product use categorization
- Outperforms[2] community-standard database, ChemSpider, for compound identification. Demonstrated using peer-reviewed publication data

Mass & Formula-based searches



Impacts

- **Curated chemistry data:** High quality curated chemistry datasets as foundation data for NCCT applications and available to other agency resources for reuse
- **Flexible Platform Design:** "Build-once, Use Many" approach provides architecture for other apps. RapidTox application already in development for internal use
- **Integration Platform**: Additional data streams can be added into the design to provide unifying integration platform for NCCT data and other agency resources

Coming Soon

- **Batch Searching:** Batch look-up and download of chemical names, CAS Numbers and DB mappings
- **Literature Handling**: Integrate to PubMed (with "Abstract Sifting" , Google Scholar, Books and Patents
- **Predicted Functional Use**: CPDat database (replacing CPCat). Includes Predicted Functional Use
- **iPhone/iPad App**: Open Data allows reuse of the data

Future Work

- **Support for ambiguous substances** – Data model will be expanded to include such ambiguous chemical substances as well as relationships within complex chemical mixtures
- **Delivery of Public Web Services** – Application programming interface and web services to allow third parties to consume and integrate these data
- **Access to real-time predictions** – Real-time prediction allowing scientists to submit one or more chemical structures and perform the predictions in real time

References

1. An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modeling, K. Mansouri, C.M. Grulke, A.M. Richard, R.S. Judson and A.J. Williams, Accepted for publication to SAR and QSAR in Environmental Research.

2. Identifying "known unknowns" using the US EPA's CompTox Chemistry Dashboard, A.D. McEachran, J.R. Sobus and A.J. Williams, Submitted for publication to Analytical and Bioanalytical Chemistry.