Bringing it all together: A web-based database for chemical and biological data to support environmental toxicology

Antony Williams[1], Chris Grulke[1], Sean Watford[1], Rebecca Jolley[1], Jeremy Dunne[1], Elizabeth Edmiston[1] and Jeff Edwards[1]

[1]U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, Research Triangle Park, NC

ACS Fall 2017
Washington, DC
August 20-24, 2017

ORCID: 0000-0002-2668-4821
Antony Williams I williams.antony@epa.gov I 919-541-1033

Problem Definition and Goals

Problem: There are many freely available data available online to support computational toxicology in environmental science, but an easy way to access available data across multiple sites is lacking.

Goals: To deliver online access via a simple to use web-based interface supporting diverse types of data associated with environmental chemistry, and specifically computational toxicology. To make the data available for ~760,000 chemical substances available as downloadable data for reuse and repurposing in other databases.

Abstract

The EPA Comptox Chemistry Dashboard is a web-based application providing access to a set of data resources provided by the National Center of Computational Toxicology. Diverse data types associated with ~760,000 chemical substances are delivered through the application. These data include bioassay screening results (associated with the ToxCast project), physicochemical and toxicological endpoints (both experimental and predicted), and consumer product and functional use data. Access to tens of millions of predicted property values associated with the chemicals on the dashboard includes detailed calculation reports for the OPERA models [1].

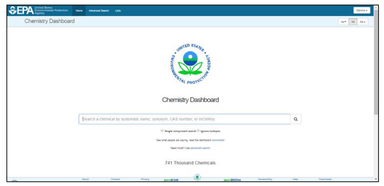
The dashboard is an integration hub for ~70 public resources using cross website linking. This helps the user navigate to other data and information for a particular chemical on other websites. The application also provides integrated literature searches for chemical related information using Google Scholar and PubMed.

Advanced searching capabilities provide molecular formula and mass-based searching to support non-targeted analysis for mass spectrometry, a key area of research at the US Environmental Protection Agency. A batch search also allows users to search using inputs of thousands of chemical names or CAS Registry Numbers and download details regarding the availability of bioassay, exposure and toxicity data.

Much of the data on the dashboard is available as Open Data via a downloads page. This allows third parties to integrate these data into their own applications and link back to the dashboard chemical pages.

The underpinning architecture has been developed in a manner allowing for the release of the dashboard as a public resource as well as for delivering applications internal to the U.S. EPA where the application is being assessed for potential applications to support risk assessment.

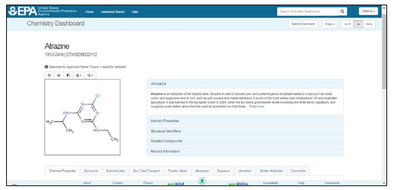
The CompTox Chemistry Dashboard



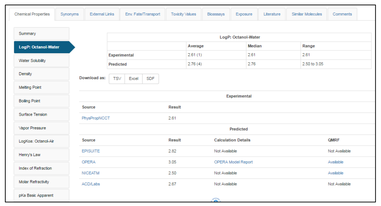
Dashboard Entry Page

The landing page of the dashboard is a simple text entry box allowing a type-ahead search for systematic, trade and trivial names, CAS Registry Numbers and InChIs.

Where possible, links are provided to related Wikipedia articles. An associated mol file is available for download to the desktop, and a summary report containing record data can be provided as a PDF file.



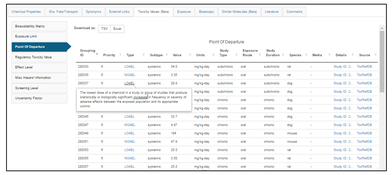
Chemical Record Page: Atrazine



Chemical Properties Panel

For records with chemical structure representations, various inherent properties (e.g. formula and mass) and predicted physicochemical properties (logP, water solubility etc.) are provided.

The Toxicity Values tab provides access to data assembled from a series of public resources including EPA data (i.e. IRIS and PPRTV reports, ToxRef DB). Data can be downloaded as TSV and Excel files.



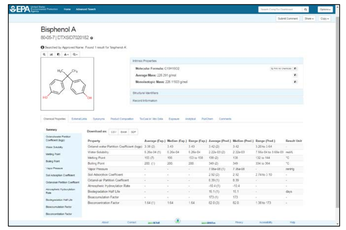
Toxicity Values Panel

Bioassays: ToxCast Data

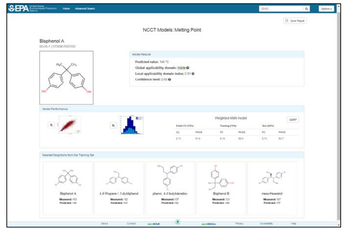
The ToxCast Bioassay data measured over the past decade is accessible via the dashboard under the Bioassays Tab. PubChem bioassays data are also available. Data can be downloaded as Excel files.

Accessing ~100 Million Predicted Properties Online

The dashboard provides access to ~760,000 chemicals and integrates curated experimental data [1] used to produce OPEn saR Application (OPERA) models. All chemicals were passed through the prediction models and detailed model reports showing global and local applicability domains and nearest neighbor results are displayed in the application. The QSAR Modeling Report Formats (QMRF) for each model are available for each predicted endpoint.



A summary of available chemical properties for Bisphenol A



The MP prediction model report for BPA – including nearest neighbors.

Predicted data from the EPA's Toxicity Estimation Software Tool (T.E.S.T) [2] is also available. This includes physicochemical and toxicity prediction models.

Future Work

- Continue to expand the data in terms of chemicals, toxicity data, additional experimental data.
- Provide access to the data via a suite of web services and fully documented API.
- Release NCCT models as interactive online prediction tools in the near future via the dashboard.

References

1. Mansouri *et al.* An automated curation procedure for addressing chemical errors and inconsistencies in public datasets in QSAR modelling, *SAR QSAR Environ Res. 2016 Nov;27(11):939-965.*
2. EPA Toxicity Estimation Software Tool (T.E.S.T.) software http://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test

Future Work

The authors would like to acknowledge specific colleagues within our center for their contributions to the development of the dashboard: Todd Martin, Nancy Baker, Richard Judson and John Wambaugh.