Accessing bulk predicted QSAR data via batch search functionality in the CompTox Chemicals Dashboard

Abstract #240

Antony J. Williams[1], Chris Grulke[1], Todd Martin[1] and Kamel Mansouri[2]
[1]Ctr. for Comput. Toxi. & Exposure, ORD, U.S. EPA; [2]NICEATM, RTP, USA
ORCID: 0000-0002-2668-4821

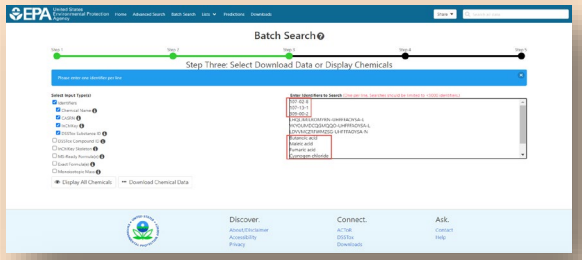
OBJECTIVES

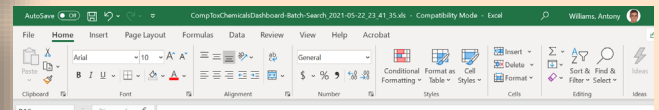
- Provide community access to experimental and predicted data for physicochemical properties, environmental fate and transport data and other QSAR outputs
- Provide batch interface to query and download data using inputs of CAS registry numbers, chemical synonyms and DSSTox substance identifiers (DTXSIDs)
- Allow for data to be intersected with multiple other data streams associated with chemicals – structure formats, formulae, masses, *in vivo* and *in vitro* toxicity data etc.

APPROACH

- CompTox Chemicals Dashboard [1] offers batch search (https://comptox.epa.gov/dashboard)
- QSAR predicted data from data streams available: EPI Suite, ACD/Labs, OPERA [2], Toxicity Estimation Software Tool (TEST) [3]
- Data include logKow, water solubility, vapor pressure, bioconcentration and bioaccumulation factors and many other properties
- Batch mode delivers data in CSV, Excel and SDF format

MAIN RESULTS





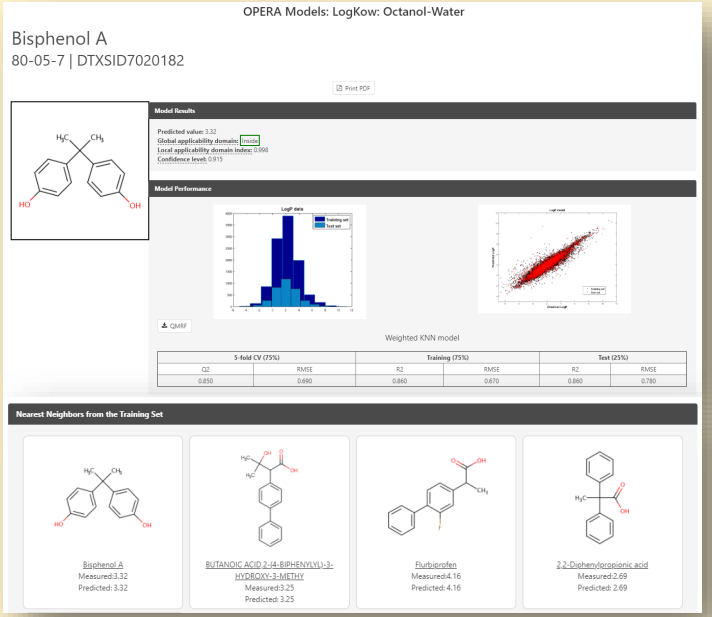
- Batch search CAS and Names
- Select data to download
- Export CSV, XLS or SDF
- FAIR data available

IMPACT

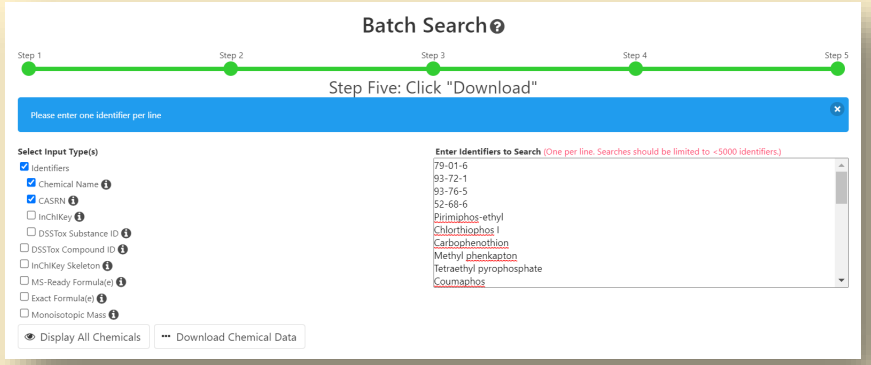
- Fast, simple access to data for >800,000 chemicals
- QSAR predicted data available, *en masse*, via a batch search [4] interface, delivers access to >25 million predicted properties
- Curated data sets used to create QSAR models available [5]

Accessing bulk predicted QSAR data via batch search functionality in the CompTox Chemicals Dashboard

- Provide community access to experimental and predicted data for physicochemical properties, environmental fate and transport data and other QSAR outputs. Available data for OPERA and TEST includes detailed calculation protocols

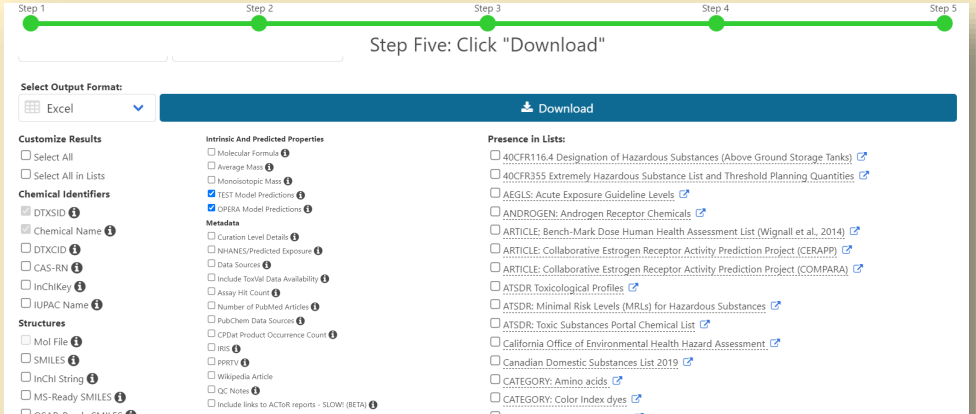


OPERA Calculation Protocol on Dashboard



Batch Search Interface on Dashboard

Provide batch interface to query and download data using inputs of CAS registry numbers and chemical synonyms

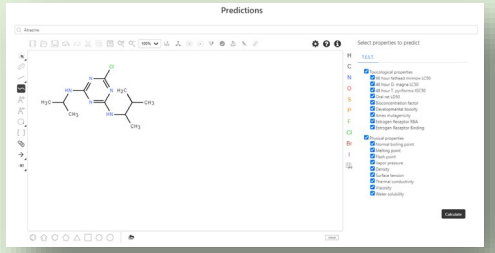


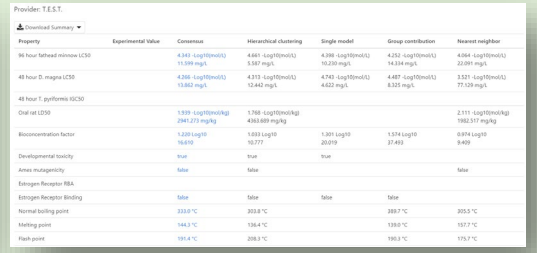
- Allow for data to be intersected with multiple other data streams associated with chemicals – structure formats, formulae, masses, *in vivo* and *in vitro* toxicity data etc.

Accessing bulk predicted QSAR data via batch search functionality in the CompTox Chemicals Dashboard

Summary

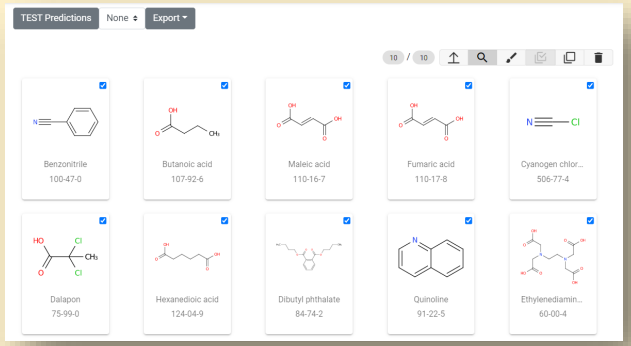
- QSAR predictions for over 800,000 chemicals, to produce over 25 million predicted values, available through a simple batch search
- REAL-TIME predictions using TEST predictions are also available [6]

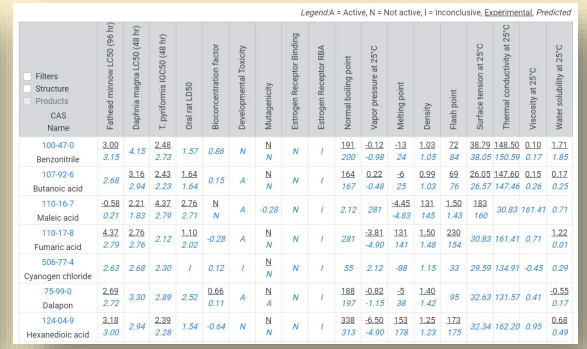




Future Plans

- Batch Prediction of data available through Dashboard





- Batch prediction proof-of-concept already in testing

References

1. Williams, A. J. et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics 9, 61 (2017)
2. Mansouri, K. et al. OPERA models for predicting physicochemical properties and environmental fate endpoints. Journal of Cheminformatics 10, 10 (2018)
3. Toicity Estimation Software Tool (TEST) : https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test
4. CN Lowe, AJ Williams, Enabling High-Throughput Searches for Multiple Chemical Data Using the Chemicals Dashboard, J. Chem. Inf. Model. 2021, 61, 2, 565–570
5. K. Mansouri, (2016) An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling, SAR and QSAR in Environmental Research, 27:11, 911-937
6. Real-time TEST predictions: https://comptox.epa.gov/dashboard/predictions/index