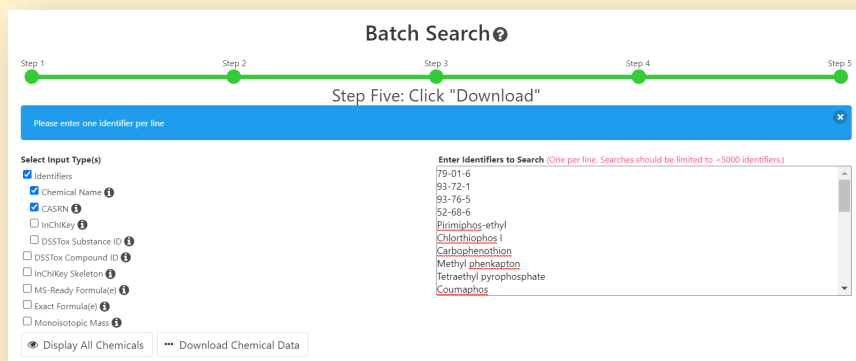


Accessing bulk predicted QSAR data via batch search functionality in the CompTox Chemicals Dashboard

- 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

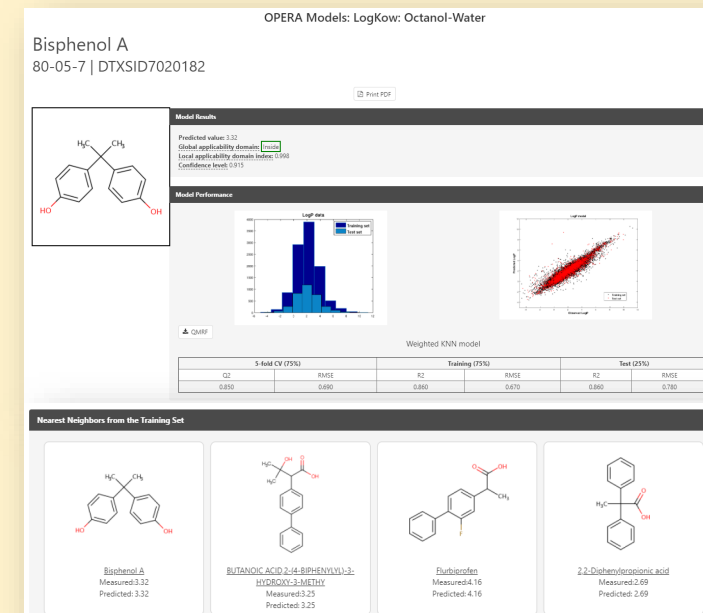


The image shows the 'Batch Search' interface in the CompTox Chemicals Dashboard. It features a progress bar at the top with five steps, and the fifth step is highlighted with the text 'Step Five: Click "Download"'. Below the progress bar is a search input field with the placeholder text 'Please enter one identifier per line'. To the left of the input field is a 'Select Input Type(s)' section with checkboxes for 'Identifiers', 'Chemical Name', 'CASRN', 'InChIKey', 'DSSTox Substance ID', 'DSSTox Compound ID', 'InChIKey Skeleton', 'MS-Ready Formulae', 'Exact Formulae', and 'Monoisotopic Mass'. To the right of the input field is a section titled 'Enter Identifiers to Search (One per line. Searches should be limited to <5000 identifiers)'. This section contains a list of identifiers: '79-01-6', '93-72-1', '93-76-5', '52-69-6', 'Phosphor-ethyl', 'Chlorophosphor', 'Carbophenothion', 'Methyl phenkapton', 'Tetraethyl pyrophosphate', and 'Coumaphos'. At the bottom left of the interface are two buttons: 'Display All Chemicals' and 'Download Chemical Data'.

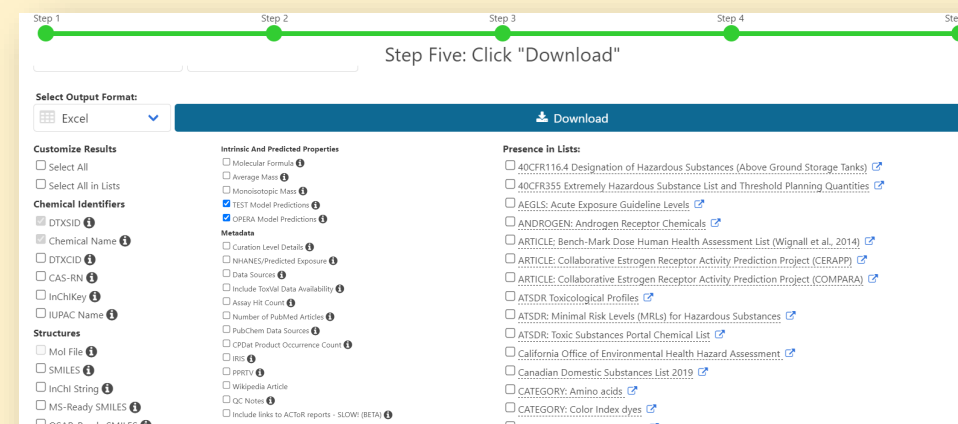
Batch Search Interface on Dashboard

- 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

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



OPERA Calculation Protocol on Dashboard

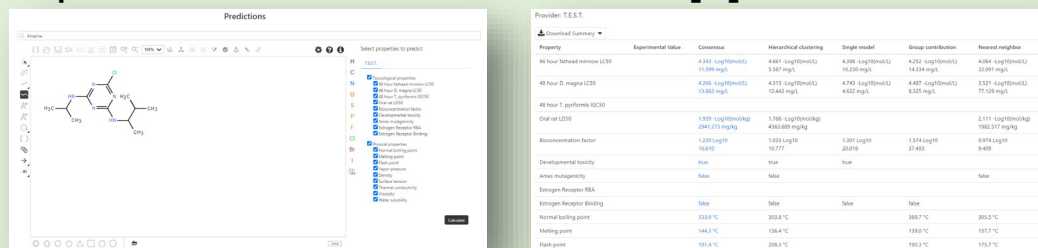


The image shows the 'Batch Search' interface in the CompTox Chemicals Dashboard, specifically the 'Step Five: Click "Download"' section. It features a progress bar at the top with five steps, and the fifth step is highlighted with the text 'Step Five: Click "Download"'. Below the progress bar is a 'Select Output Format:' section with a dropdown menu set to 'Excel'. To the right of the dropdown is a 'Download' button. Below these are three sections: 'Customize Results' with checkboxes for 'Select All', 'Select All in Lists', 'Chemical Identifiers', 'DTXSID', 'Chemical Name', 'DTXGID', 'CAS-RN', 'InChIKey', and 'IUPAC Name'; 'Structures' with checkboxes for 'Mol File', 'SMILES', 'InChI String', 'MS-Ready SMILES', and 'InChIKey'; and 'Intrinsic And Predicted Properties' with checkboxes for 'Molecular Formula', 'Average Mass', 'Monoisotopic Mass', 'TEST Model Predictions', 'OPERA Model Predictions', 'Correlation Level Details', 'NHANES Predicted Exposure', 'Data Sources', 'Include ToxVal Data Availability', 'Assay Hit Count', 'Number of Published Articles', 'PubChem Data Sources', and 'Other Product Occurrence Count'. To the right of these sections is a 'Presence in Lists:' section with a list of links to various databases and lists, including '40CFR116.4 Designation of Hazardous Substances (Above Ground Storage Tanks)', '40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities', 'AEGIS: Acute Exposure Guideline Levels', 'ANDROGEN: Androgen Receptor Chemicals', 'ARTICLE: Bench-Mark Dose Human Health Assessment List (Wignall et al., 2014)', 'ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP)', 'ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (COMPARA)', 'ATSDR Toxicological Profiles', 'ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances', 'ATSDR: Toxic Substances Portal Chemical List', 'California Office of Environmental Health Hazard Assessment', 'Canadian Domestic Substances List 2019', 'CATEGORY: Amino acids', and 'CATEGORY: Color Index dyes'.

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

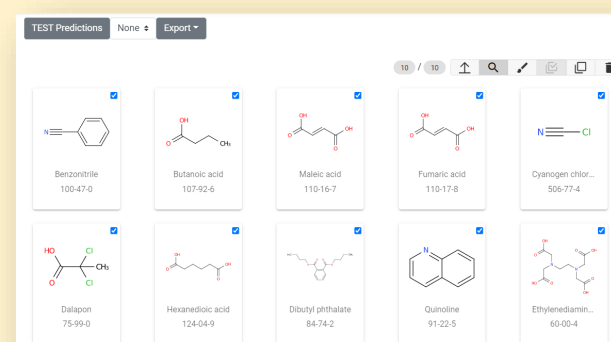
Summary

- 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



Legend: A = Active, N = Not active, I = Inconclusive, Experimental, Predicted

CAS Name	Fathead minnow LC50 (96 hr)	Daphnia magna LC50 (48 hr)	T. pyriformis LC50 (48 hr)	Oral rat LD50	Bioconcentration factor	Developmental Toxicity	Mutagenicity	Estrogen Receptor Binding	Estrogen Receptor Binding	Normal boiling point	Vapor pressure at 25 °C	Melting point	Density	Flash point	Surface tension at 25 °C	Thermal conductivity at 25 °C	Viscosity at 25 °C	Water solubility at 25 °C
100-47-0	3.00	2.48	1.57	0.85	N	N	N	N	N	191	-0.12	-13	1.03	72	38.79	148.50	0.10	1.71
Benzonitrile	3.15	2.73	1.57	0.85	N	N	N	N	N	200	-0.98	24	1.05	84	38.03	150.89	0.17	1.85
107-92-6	2.68	2.43	1.64	0.15	A	N	N	N	N	167	-0.48	25	1.03	76	26.57	147.46	0.26	0.25
Butanoic acid	0.58	2.21	4.37	2.76	N	A	-0.28	N	I	212	281	-4.45	131	1.50	193	30.83	161.41	0.71
110-16-7	0.21	1.63	2.79	2.71	N	A	-0.28	N	I	212	281	-4.45	131	1.50	193	30.83	161.41	0.71
Maleic acid	4.37	2.76	2.12	1.10	-0.28	A	N	N	N	281	-3.81	131	1.50	230	30.83	161.41	0.71	1.22
Fumaric acid	2.79	2.76	2.12	2.02	-0.28	A	N	N	N	281	-4.90	141	1.48	154	30.83	161.41	0.71	0.01
506-77-4	2.63	2.68	2.30	1.12	1	N	N	N	N	55	2.12	-88	1.15	33	29.59	134.91	-0.45	0.29
Cyanogen chloride	75-99-0	2.63	2.68	2.30	1.12	1	N	N	N	55	2.12	-88	1.15	33	29.59	134.91	-0.45	0.29
75-99-0	2.63	2.68	2.30	1.12	1	N	N	N	N	55	2.12	-88	1.15	33	29.59	134.91	-0.45	0.29
Dalapon	2.72	2.80	2.89	2.52	0.11	A	N	N	N	188	-0.82	-5	1.40	95	32.63	131.57	0.41	0.17
124-04-9	3.18	2.29	1.54	-0.64	N	N	N	N	N	338	-6.50	153	1.25	173	32.34	162.20	0.95	0.68
Hexanedioic acid	3.00	2.94	2.28	1.54	-0.64	N	N	N	N	319	-4.90	179	1.23	175	32.34	162.20	0.95	0.49

- 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>