http://www.orcid.org/0000-0002-2668-4821

Utilizing the US-EPA CompTox Chemicals Dashboard to deliver public access to a human **volatilome** subset of data

Antony Williams[1] and Joachim D. Pleil[2]

1) National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC

2) National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC

Breath research and analytical chemistry

- As a community, we often run standard GC-MS analyses on breath volatiles (or LC-MS for condensates and aerosols)
- We compare groups of samples to each other to see what is different (sick/not sick, exposed/not exposed, before/after intervention).
- We use the differences in chemicals or chemical features to infer how the body responded.
- These turn into “pre-clinical markers”

Identification of chemical features

- In non-targeted analysis, we often do not know the identity of an analytical feature besides retention time (RT) and mass spectrum.

- There are two schools of thought:
  - **We don't care**; as long as the RT and mass fragments are consistent, we can develop a case-control pattern between groups.
  - **We do care**; we would like to also understand the biochemistry that results in group differences for which we need unambiguous id's.

Identification of chemical features

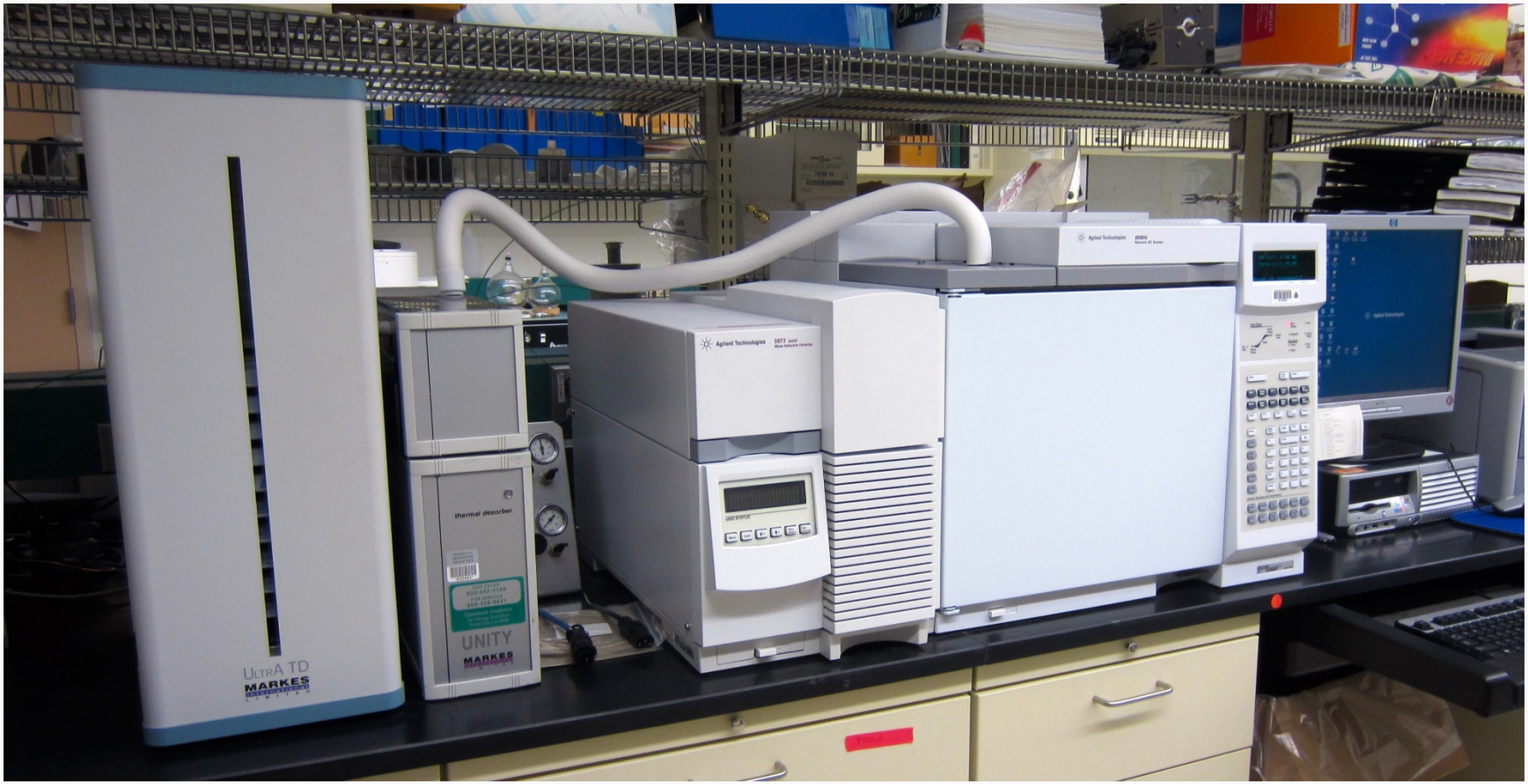
- Both approaches have value. The second one is often not possible.
- There are many of thousands of environmental and biological chemicals in human media (exposome).
- A large portion of these have never been unambiguously identified.
- We rely on library searches (NIST) and other software tools to make tentative assignments.

These external tools are agnostic

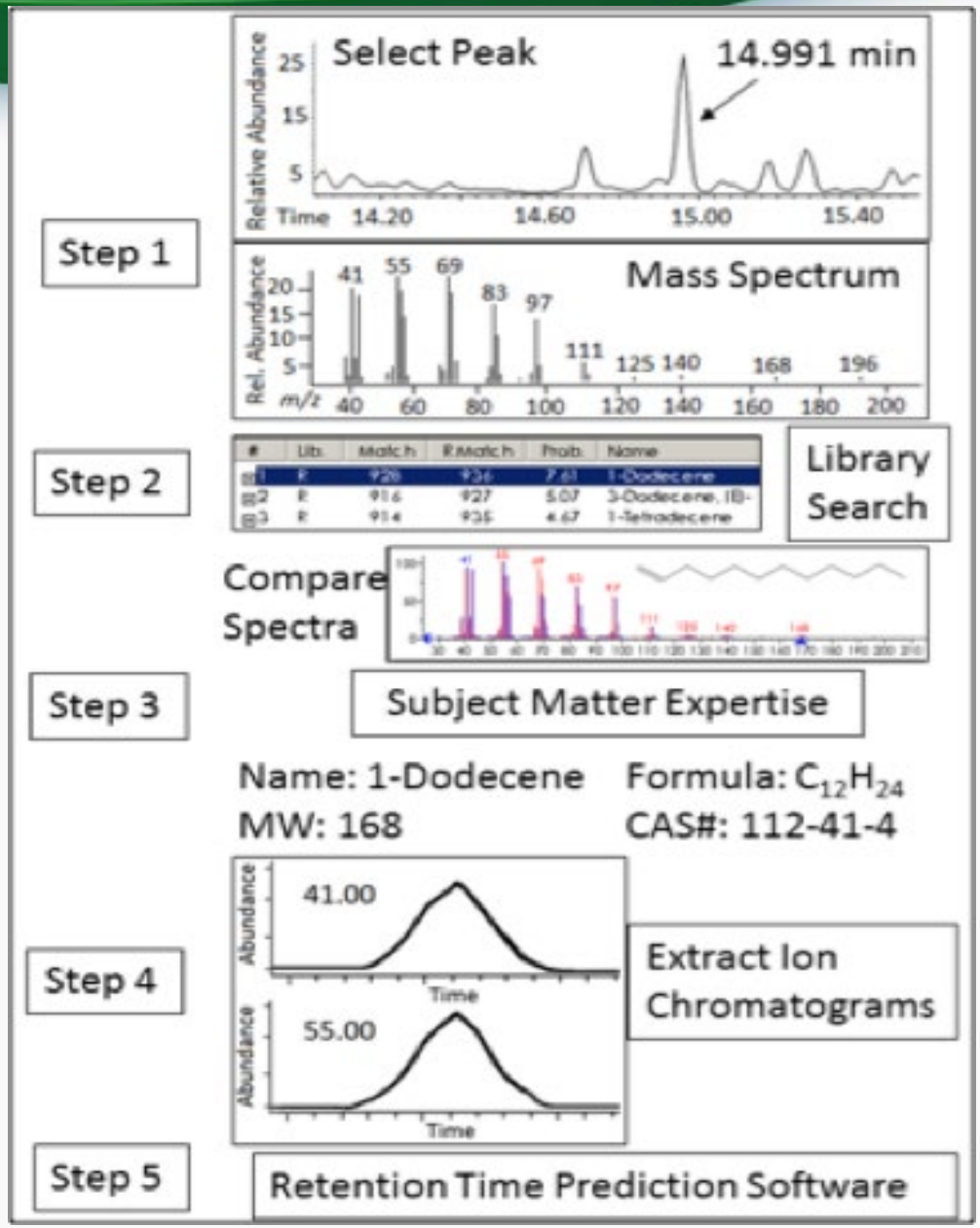
- We can get a list of 30 or more candidate compounds from a simple NIST search for one feature.
- Some may be “impossible” for breath.
- Some may be “plausible” for breath.
- However, it requires subject matter expertise and a tiered workflow to narrow down identifications – see for example:

Wallace MAG, Pleil JD, Oliver KD, Whitaker DA, Mentese S, Fent KW, and Horn GP, 2019. Non-targeted GC-MS analysis of exhaled breath samples: Exploring human biomarkers of exogenous exposure and endogenous response from professional firefighting activity. Journal of Toxicology and Environmental Health, Part A 82(4): 244-260.

For example: standard TD-GC-MS



Workflow example:



Example: Molecular (integer mass) 73

Tentative ID	Chem Formula	Int. Mass	Exact mass
Methyl isothiocyanate	C2H3NS	73	72.9986201
Methyl thiocyanate	C2H3NS	73	72.9986201
Ethene, nitro-	C2H3NO2	73	73.01637834
Acetoxime	C3H7NO	73	73.05276385
N,N-Dimethylformamide	C3H7NO	73	73.05276385
N-Methylacetamide	C3H7NO	73	73.05276385
Propanamide	C3H7NO	73	73.05276385
Formamide, N-ethyl-	C3H7NO	73	73.05276385
Formamide, N,N-dimethyl-, compd. with sulfur trioxide (1:1)	C3H7NO	73	73.05276385
Formic acid, compd. with N,N-dimethylformamide (1:1)	C3H7NO	73	73.05276385
1-Methylguanidine	C2H7N3	73	73.06399724
Guanidine, methyl-, sulfate (2:1)	C2H7N3	73	73.06399724
2-Butanol, aluminum salt	C4H9O	73	73.06533991
2-Propanol, 2-methyl-, aluminum salt	C4H9O	73	73.06533991
1-Butanol, antimony(3+) salt	C4H9O	73	73.06533991
1-Butanol, titanium(4+) salt	C4H9O	73	73.06533991
1-Propanol, 2-methyl-, titanium(4+) salt	C4H9O	73	73.06533991
1-Butanol, potassium salt	C4H9O	73	73.06533991
2-Methyl-2-propanol, potassium salt	C4H9O	73	73.06533991
Lithium 2-methylpropan-2-olate	C4H9O	73	73.06533991
1-Butanol, sodium salt	C4H9O	73	73.06533991
2-Propanol, 2-methyl-, sodium salt	C4H9O	73	73.06533991
Butylamine	C4H11N	73	73.08914936
Diethylamine	C4H11N	73	73.08914936
2-Butanamine	C4H11N	73	73.08914936
tert-Butylamine	C4H11N	73	73.08914936
Isobutylamine	C4H11N	73	73.08914936
N,N-Dimethylethylamine	C4H11N	73	73.08914936
Diethylamine hydrochloride	C4H11N	73	73.08914936
2-Propanamine, 2-methyl-, hydrochloride	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, perchlorate	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, sulfate	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, phosphate (1:1)	C4H11N	73	73.08914936
1-Butanamine, acetate	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, acetate	C4H11N	73	73.08914936
2-Propanamine, 2-methyl-, sulfate (2:1)	C4H11N	73	73.08914936

Example: Molecular (integer mass) 73

Tentative ID	Chem Formula	Int. Mass	Exact mass
Methyl isothiocyanate	C2H3NS	73	72.9986201
Methyl thiocyanate	C2H3NS	73	72.9986201
Ethene, nitro-	C2H3NO2	73	73.01637834
Acetoxime	C3H7NO	73	73.05276385
N,N-Dimethylformamide	C3H7NO	73	73.05276385
N-Methylacetamide	C3H7NO	73	73.05276385
Propanamide	C3H7NO	73	73.05276385
Formamide, N-ethyl-	C3H7NO	73	73.05276385
Formamide, N,N-dimethyl-, compd. with sulfur trioxide (1:1)	C3H7NO	73	73.05276385
Formic acid, compd. with N,N-dimethylformamide (1:1)	C3H7NO	73	73.05276385
1-Methylguanidine	C2H7N3	73	73.06399724
Guanidine, methyl-, sulfate (2:1)	C2H7N3	73	73.06399724
2-Butanol, aluminum salt	C4H9O	73	73.06533991
2-Propanol, 2-methyl-, aluminum salt	C4H9O	73	73.06533991
1-Butanol, antimony(3+) salt	C4H9O	73	73.06533991
1-Butanol, titanium(4+) salt	C4H9O	73	73.06533991
1-Propanol, 2-methyl-, titanium(4+) salt	C4H9O	73	73.06533991
1-Butanol, potassium salt	C4H9O	73	73.06533991
2-Methyl-2-propanol, potassium salt	C4H9O	73	73.06533991
Lithium 2-methylpropan-2-olate	C4H9O	73	73.06533991
1-Butanol, sodium salt	C4H9O	73	73.06533991
2-Propanol, 2-methyl-, sodium salt	C4H9O	73	73.06533991
Butylamine	C4H11N	73	73.08914936
Diethylamine	C4H11N	73	73.08914936
2-Butanamine	C4H11N	73	73.08914936
tert-Butylamine	C4H11N	73	73.08914936
Isobutylamine	C4H11N	73	73.08914936
N,N-Dimethylethylamine	C4H11N	73	73.08914936
Diethylamine hydrochloride	C4H11N	73	73.08914936
2-Propanamine, 2-methyl-, hydrochloride	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, perchlorate	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, sulfate	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, phosphate (1:1)	C4H11N	73	73.08914936
1-Butanamine, acetate	C4H11N	73	73.08914936
Ethanamine, N-ethyl-, acetate	C4H11N	73	73.08914936
2-Propanamine, 2-methyl-, sulfate (2:1)	C4H11N	73	73.08914936

Example: Molecular (integer) mass 86

Tentative ID	Chem Formula	Int. Mass	Exact mass
Chlorodifluoromethane	CHClF2	86	85.97348415
Methoxy, trifluoro-	CHF3O	86	85.99794931
1,3-Dioxol-2-one	C3H2O3	86	86.00039392
Sodium methacrylate	C4H5NaO2	86	86.03677943
beta-Butyrolactone	C4H6O2	86	86.03677943
4-Butyrolactone	C4H6O2	86	86.03677943
dl-Diepoxybutane	C4H6O2	86	86.03677943
Vinyl acetate	C4H6O2	86	86.03677943
Butanedial	C4H6O2	86	86.03677943
2,3-Butanedione	C4H6O2	86	86.03677943
2-Butyne-1,4-diol	C4H6O2	86	86.03677943
Methyl acrylate	C4H6O2	86	86.03677943
Methacrylic acid	C4H6O2	86	86.03677943
2-Butenoic acid	C4H6O2	86	86.03677943
Cyclopropane carboxylic acid	C4H6O2	86	86.03677943
2,2'-Bioxirane	C4H6O2	86	86.03677943
Formic acid, 2-propenyl ester	C4H6O2	86	86.03677943
1,4-Dioxin, 2,3-dihydro-	C4H6O2	86	86.03677943
Butanal, 2-oxo-	C4H6O2	86	86.03677943
Ammonium methacrylate	C4H9NO2	86	86.03677943
Vinyl acetate ethylene copolymer	C6H10O2	86	86.03677943
Vinylidene chloride/methylacrylate copolymer	C6H8Cl2O2	86	86.03677943
Lead(II) methacrylate	C8H10O4Pb	86	86.03677943
Zinc dimethacrylate	C8H10O4Zn	86	86.03677943
Ethylene urea	C3H6N2O	86	86.04801282
N-Nitrosoazetidine	C3H6N2O	86	86.04801282
Aluminum, diethylhydro-	C4H11Al	86	86.06761398
3-Methylbutanal	C5H10O	86	86.07316494
Pentanal	C5H10O	86	86.07316494
2-Methylbutanal	C5H10O	86	86.07316494
3-Pentanone	C5H10O	86	86.07316494
2-Pentanone	C5H10O	86	86.07316494
3-Methyl-2-butanone	C5H10O	86	86.07316494
3-Methyl-2-buten-1-ol	C5H10O	86	86.07316494
2-Methyltetrahydrofuran	C5H10O	86	86.07316494
Cyclopentanol	C5H10O	86	86.07316494
2-Methyl-3-buten-2-ol	C5H10O	86	86.07316494
3-Methylbut-3-en-1-ol	C5H10O	86	86.07316494
Tetrahydropyran	C5H10O	86	86.07316494
1-Propene, 3-ethoxy-	C5H10O	86	86.07316494
Propanal, 2,2-dimethyl-	C5H10O	86	86.07316494
Propane, 1-(ethenyloxy)-	C5H10O	86	86.07316494
Propane, 2-(ethenyloxy)-	C5H10O	86	86.07316494
2-Buten-1-ol, 2-methyl-	C5H10O	86	86.07316494
Piperazine	C4H10N2	86	86.08439833
Diaziridine, 3-ethyl-3-methyl-	C4H10N2	86	86.08439833
Piperazine, monohydrochloride	C4H11ClN2	86	86.08439833
Piperazine dihydrochloride	C4H12Cl2N2	86	86.08439833
Carbamodithioic acid, compd. with piperazine	C5H13N3S2	86	86.08439833
Hexane	C6H14	86	86.10955045
2,2-Dimethylbutane	C6H14	86	86.10955045
2,3-Dimethylbutane	C6H14	86	86.10955045
2-Methylpentane	C6H14	86	86.10955045
3-Methylpentane	C6H14	86	86.10955045

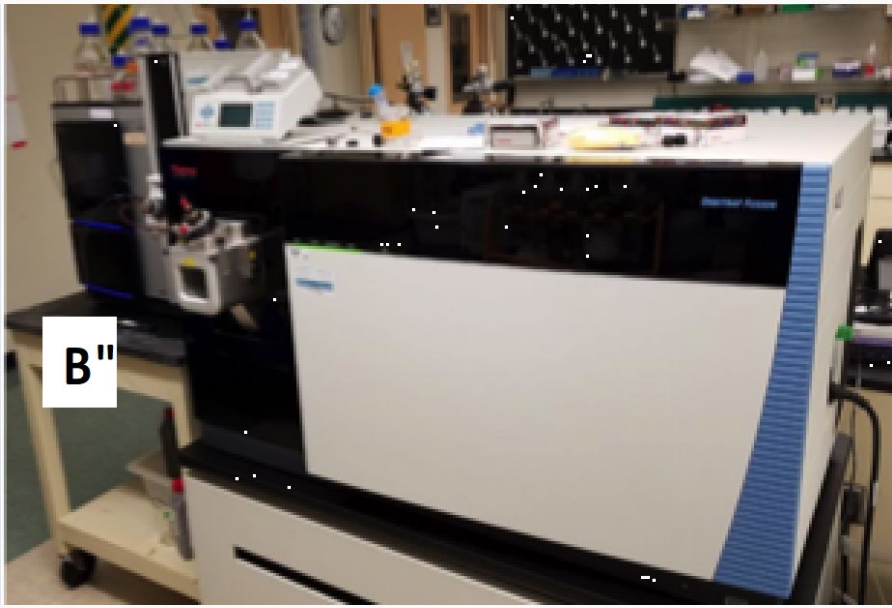
Example: Molecular (integer) mass 86

Tentative ID	Chem Formula	Int. Mass	Exact mass
Chlorodifluoromethane	CHClF2	86	85.97348415
Methoxy, trifluoro-	CHF3O	86	85.99794931
1,3-Dioxol-2-one	C3H2O3	86	86.00039392
Sodium methacrylate	C4H5NaO2	86	86.03677943
beta-Butyrolactone	C4H6O2	86	86.03677943
4-Butyrolactone	C4H6O2	86	86.03677943
dl-Diepoxybutane	C4H6O2	86	86.03677943
Vinyl acetate	C4H6O2	86	86.03677943
Butanedial	C4H6O2	86	86.03677943
2,3-Butanedione	C4H6O2	86	86.03677943
2-Butyne-1,4-diol	C4H6O2	86	86.03677943
Methyl acrylate	C4H6O2	86	86.03677943
Methacrylic acid	C4H6O2	86	86.03677943
2-Butenoic acid	C4H6O2	86	86.03677943
Cyclopropane carboxylic acid	C4H6O2	86	86.03677943
2,2'-Bioxirane	C4H6O2	86	86.03677943
Formic acid, 2-propenyl ester	C4H6O2	86	86.03677943
1,4-Dioxin, 2,3-dihydro-	C4H6O2	86	86.03677943
Butanal, 2-oxo-	C4H6O2	86	86.03677943
Ammonium methacrylate	C4H9NO2	86	86.03677943
Vinyl acetate ethylene copolymer	C6H10O2	86	86.03677943
Vinylidene chloride/methylacrylate copolymer	C6H8Cl2O2	86	86.03677943
Lead(II) methacrylate	C8H10O4Pb	86	86.03677943
Zinc dimethacrylate	C8H10O4Zn	86	86.03677943
Ethylene urea	C3H6N2O	86	86.04801282
N-Nitrosoazetidine	C3H6N2O	86	86.04801282
Aluminum, diethylhydro-	C4H11Al	86	86.05761398
3-Methylbutanal	C5H10O	86	86.07316494
Pentanal	C5H10O	86	86.07316494
2-Methylbutanal	C5H10O	86	86.07316494
3-Pentanone	C5H10O	86	86.07316494
2-Pentanone	C5H10O	86	86.07316494
3-Methyl-2-butanone	C5H10O	86	86.07316494
3-Methyl-2-buten-1-ol	C5H10O	86	86.07316494
2-Methyltetrahydrofuran	C5H10O	86	86.07316494
Cyclopentanol	C5H10O	86	86.07316494
2-Methyl-3-buten-2-ol	C5H10O	86	86.07316494
3-Methylbut-3-en-1-ol	C5H10O	86	86.07316494
Tetrahydropyran	C5H10O	86	86.07316494
1-Propene, 3-ethoxy-	C5H10O	86	86.07316494
Propanal, 2,2-dimethyl-	C5H10O	86	86.07316494
Propane, 1-(ethenyloxy)-	C5H10O	86	86.07316494
Propane, 2-(ethenyloxy)-	C5H10O	86	86.07316494
2-Buten-1-ol, 2-methyl-	C5H10O	86	86.07316494
Piperazine	C4H10N2	86	86.08439833
Diaziridine, 3-ethyl-3-methyl-	C4H10N2	86	86.08439833
Piperazine, monohydrochloride	C4H11ClN2	86	86.08439833
Piperazine dihydrochloride	C4H12Cl2N2	86	86.08439833
Carbamodithioic acid, compd. with piperazine	C5H13N3S2	86	86.08439833
Hexane	C6H14	86	86.10955045
2,2-Dimethylbutane	C6H14	86	86.10955045
2,3-Dimethylbutane	C6H14	86	86.10955045
2-Methylpentane	C6H14	86	86.10955045
3-Methylpentane	C6H14	86	86.10955045

Other instruments: HR-LC-MS

Similarly, we can use external searches to try to identify semi- and non-volatile compounds from breath condensates and aerosols samples

a) Agilent LC Q-Tof



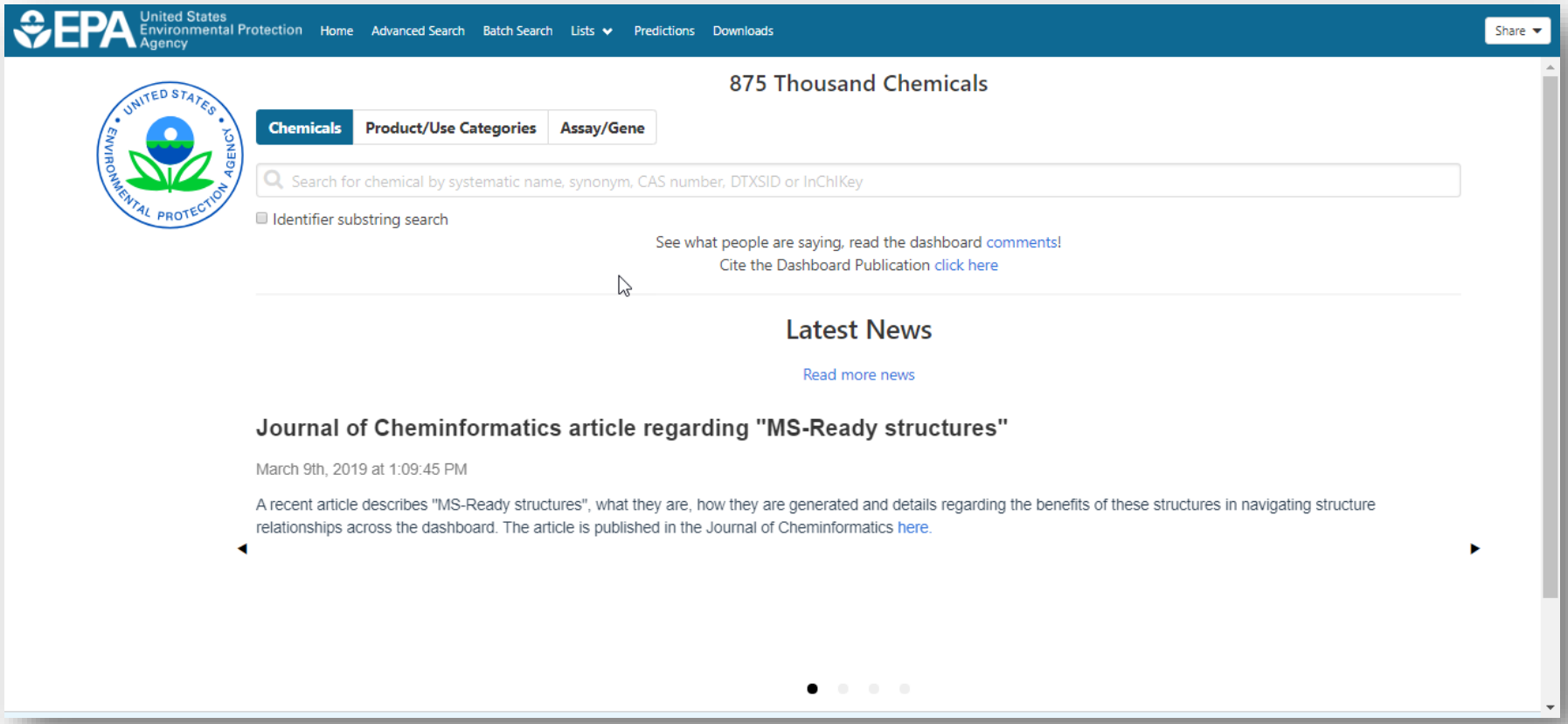
b) Agilent LC-Orbitrap

Value of CompTox Chemicals Dashboard

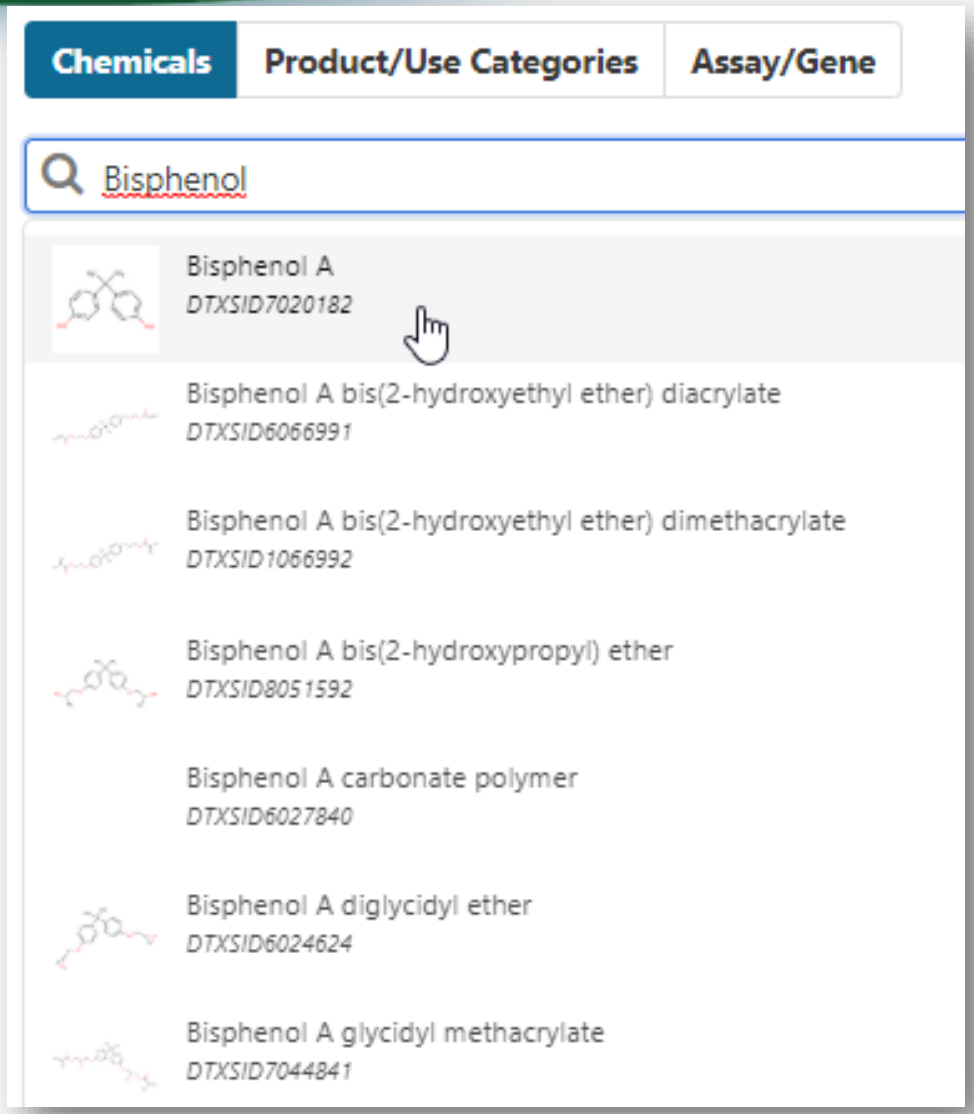
- In the following, the various functions of the EPA CompTox Chemicals Dashboard are described.
- This is a “high-level” description of a complex system; a learning curve is to be expected.
- The dashboard represents the most recent advances in chemical database search and consolidation.
- We are working on a tutorial for JBR with specific instructions for breath applications.

What is the CompTox Chemicals Dashboard?

- Freely available web-based database containing >875,000 chemical substances

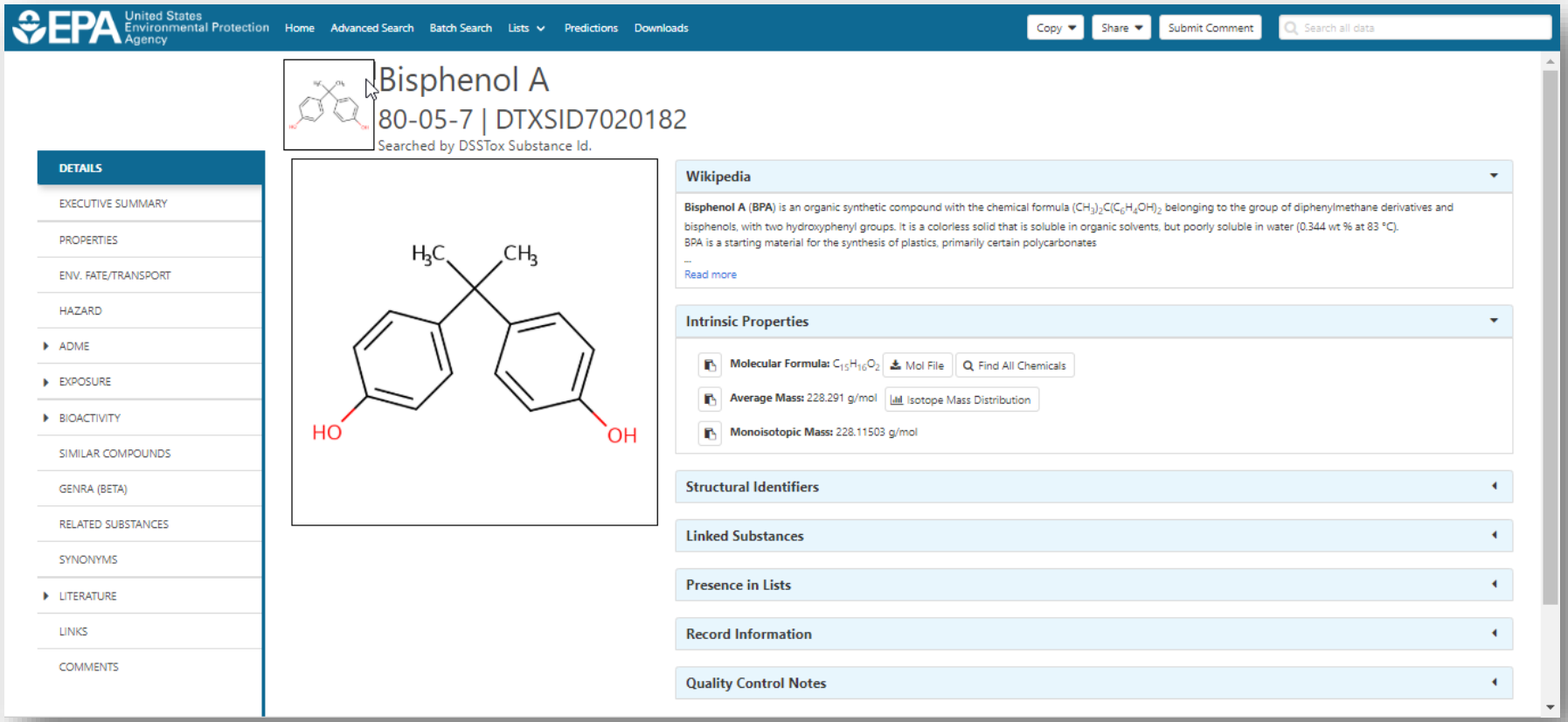


BASIC Search for chemicals



What is the CompTox Chemicals Dashboard?

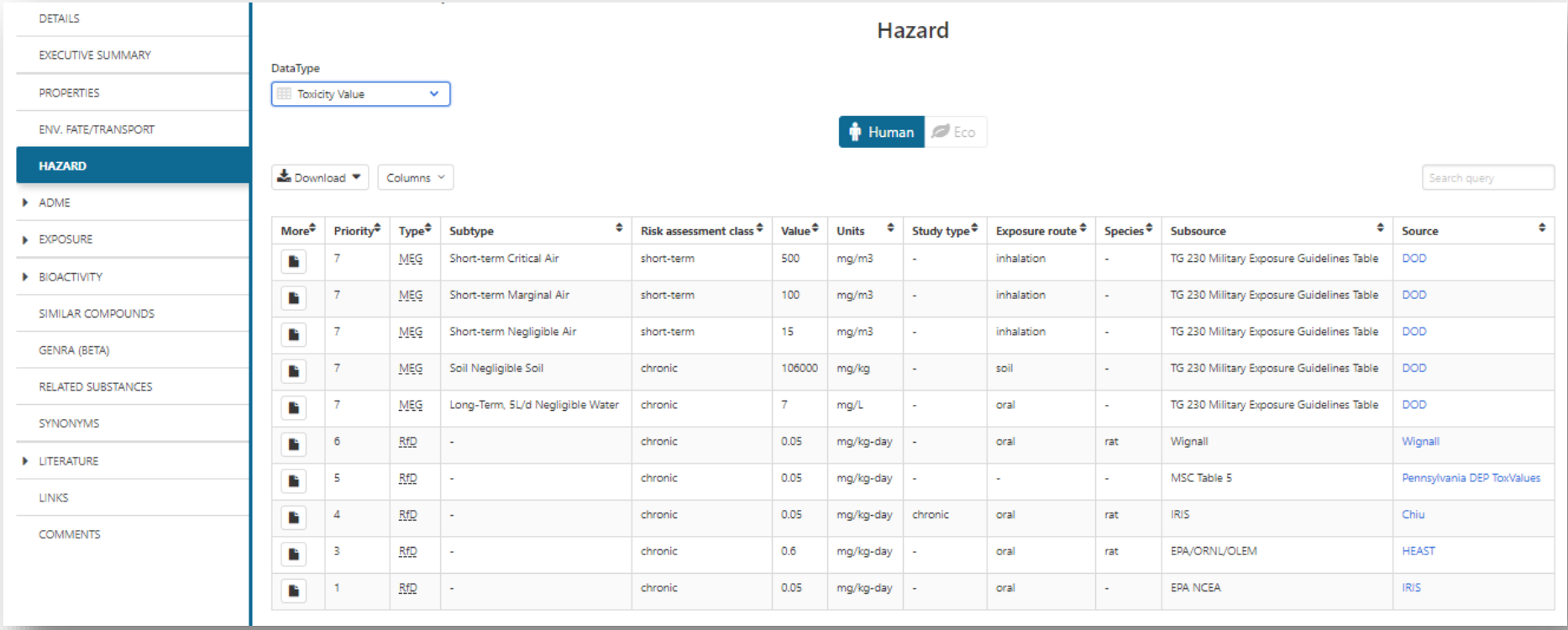
- High quality curated data



What is the CompTox Chemicals Dashboard?

EPA United States Environmental Protection Agency

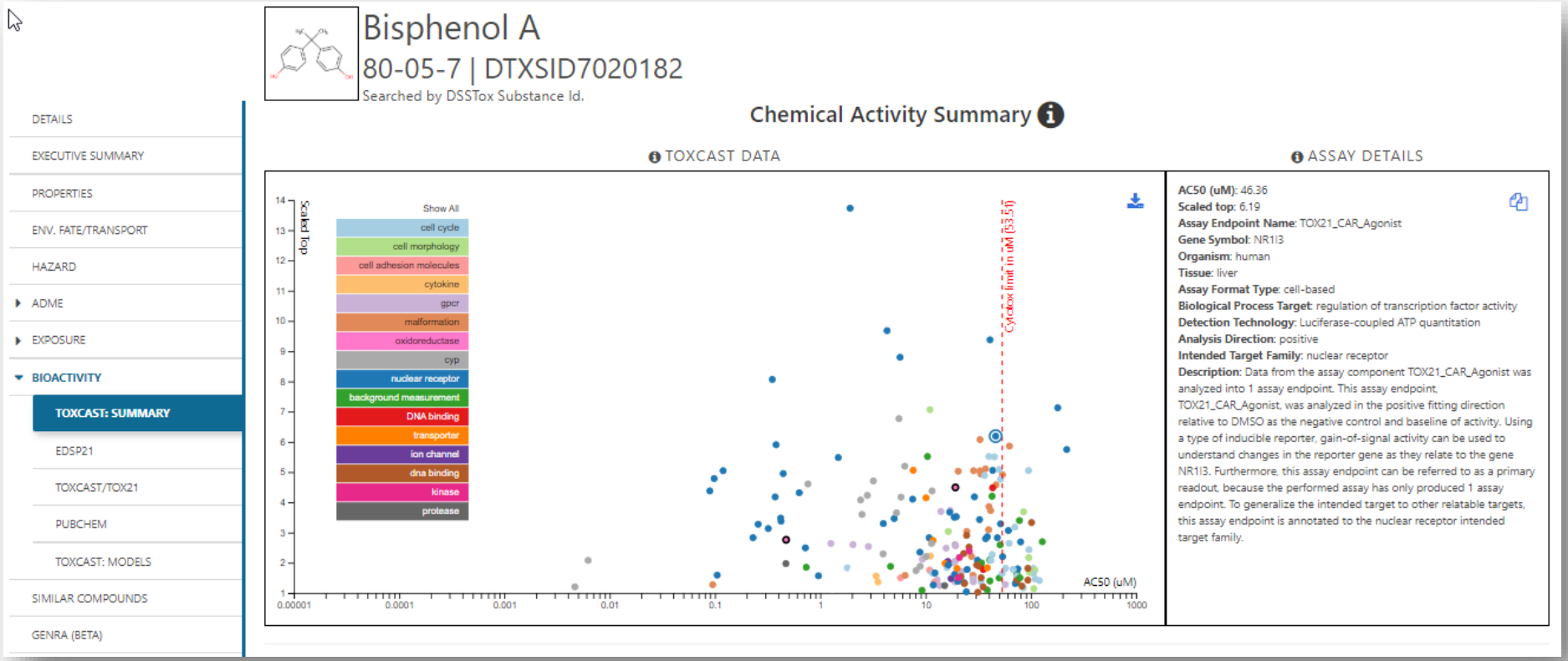
- High quality curated data – **Hazard data**



More	Priority	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	7	MEG	Short-term Critical Air	short-term	500	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Marginal Air	short-term	100	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Negligible Air	short-term	15	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Soil Negligible Soil	chronic	106000	mg/kg	-	soil	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Long-Term, 5L/d Negligible Water	chronic	7	mg/L	-	oral	-	TG 230 Military Exposure Guidelines Table	DOD
	6	RfD	-	chronic	0.05	mg/kg-day	-	oral	rat	Wignall	Wignall
	5	RfD	-	chronic	0.05	mg/kg-day	-	-	-	MSC Table 5	Pennsylvania DEP ToxValues
	4	RfD	-	chronic	0.05	mg/kg-day	chronic	oral	rat	IRIS	Chiu
	3	RfD	-	chronic	0.6	mg/kg-day	-	oral	rat	EPA/ORNL/OLEM	HEAST
	1	RfD	-	chronic	0.05	mg/kg-day	-	oral	-	EPA NCEA	IRIS

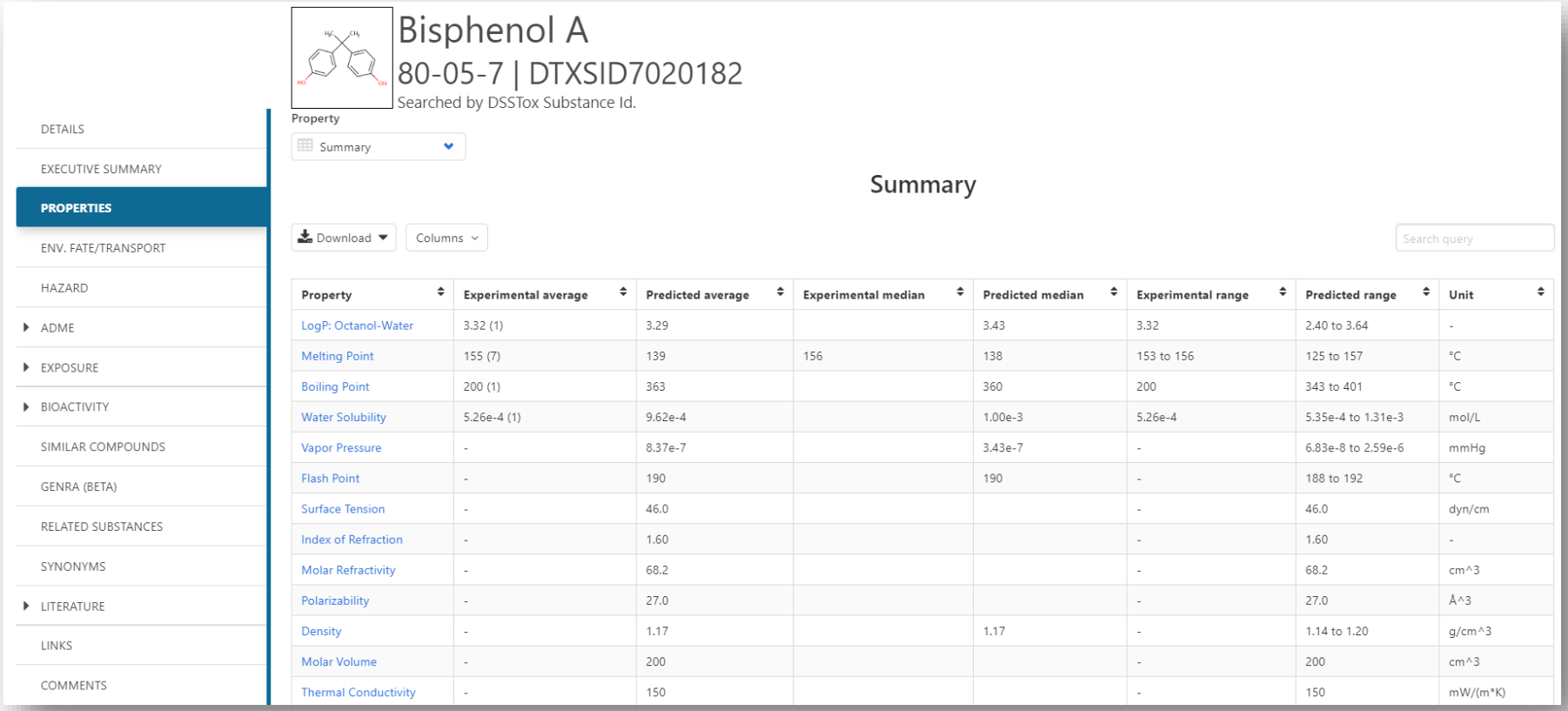
What is the CompTox Chemicals Dashboard?

- High quality curated data – Hazard data, **Bioactivities**,



What is the CompTox Chemicals Dashboard?

- High quality curated data – Hazard data, Bioactivities, **Experimental and Predicted properties**



DETAILS
EXECUTIVE SUMMARY
PROPERTIES
ENV. FATE/TRANSPORT
HAZARD
ADME
EXPOSURE
BIOACTIVITY
SIMILAR COMPOUNDS
GENRA (BETA)
RELATED SUBSTANCES
SYNONYMS
LITERATURE
LINKS
COMMENTS

Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Property

Summary

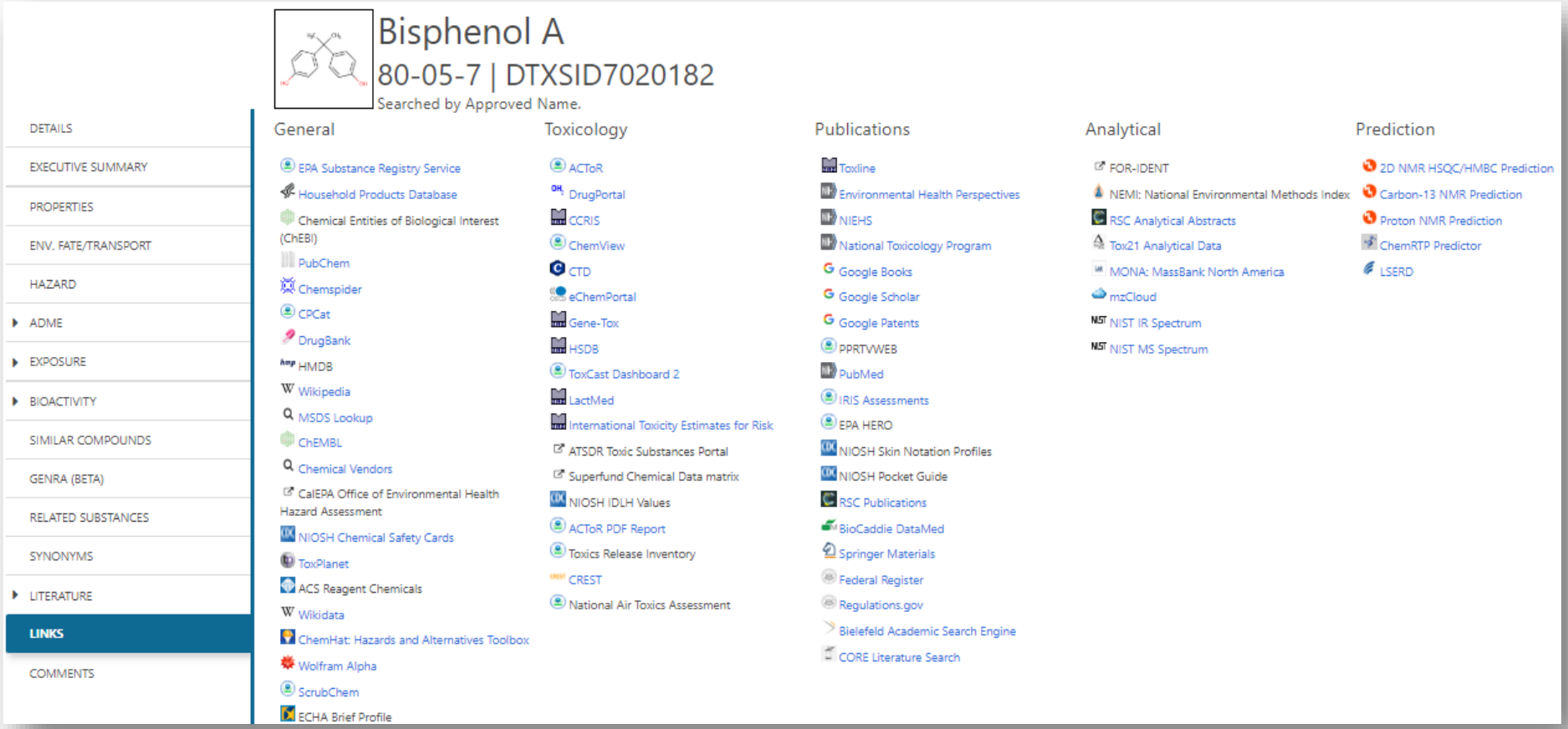
Summary

Download | Columns | Search query

Property	Experimental average	Predicted average	Experimental median	Predicted median	Experimental range	Predicted range	Unit
LogP: Octanol-Water	3.32 (1)	3.29		3.43	3.32	2.40 to 3.64	-
Melting Point	155 (7)	139	156	138	153 to 156	125 to 157	°C
Boiling Point	200 (1)	363		360	200	343 to 401	°C
Water Solubility	5.26e-4 (1)	9.62e-4		1.00e-3	5.26e-4	5.35e-4 to 1.31e-3	mol/L
Vapor Pressure	-	8.37e-7		3.43e-7	-	6.83e-8 to 2.59e-6	mmHg
Flash Point	-	190		190	-	188 to 192	°C
Surface Tension	-	46.0			-	46.0	dyn/cm
Index of Refraction	-	1.60			-	1.60	-
Molar Refractivity	-	68.2			-	68.2	cm^3
Polarizability	-	27.0			-	27.0	Å^3
Density	-	1.17		1.17	-	1.14 to 1.20	g/cm^3
Molar Volume	-	200			-	200	cm^3
Thermal Conductivity	-	150			-	150	mW/(m*K)

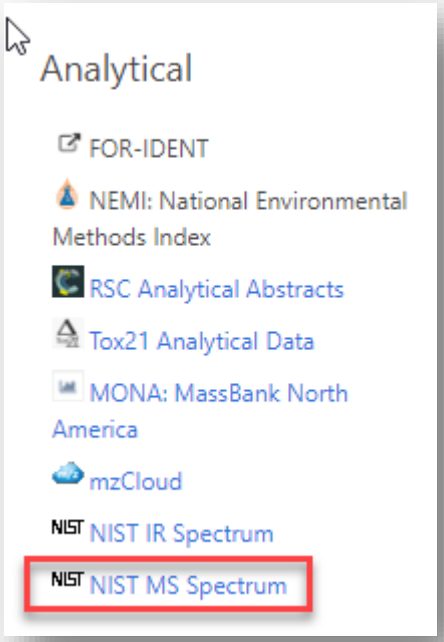
What is the CompTox Chemicals Dashboard?

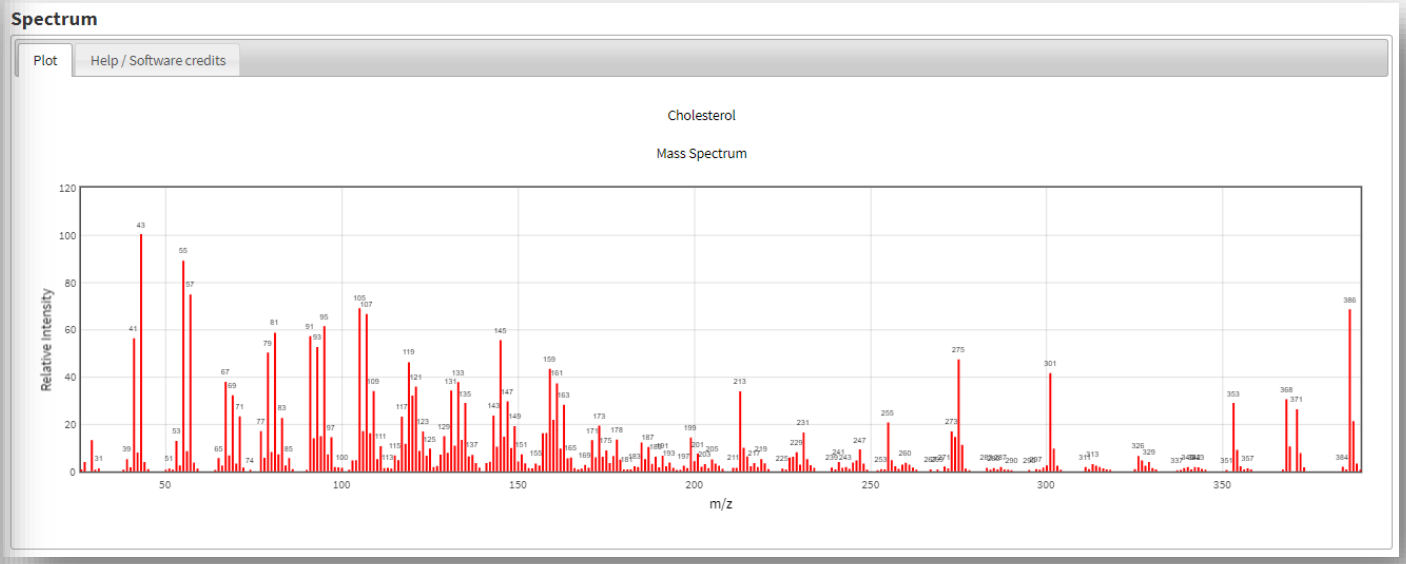
- High quality curated data – Hazard data, Bioactivities, Experimental and Predicted properties, **Links**



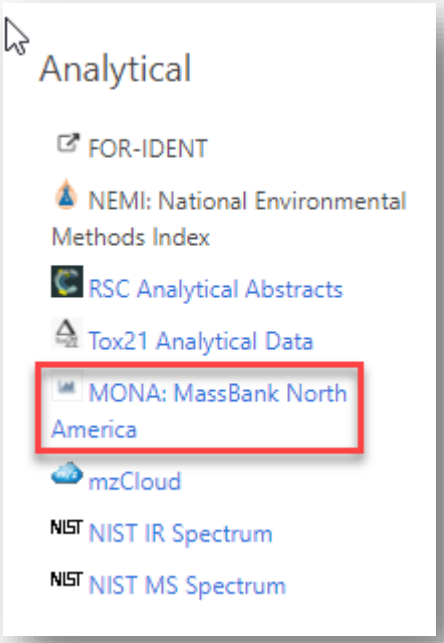
NIST WebBook
https://webbook.nist.gov/chemistry/

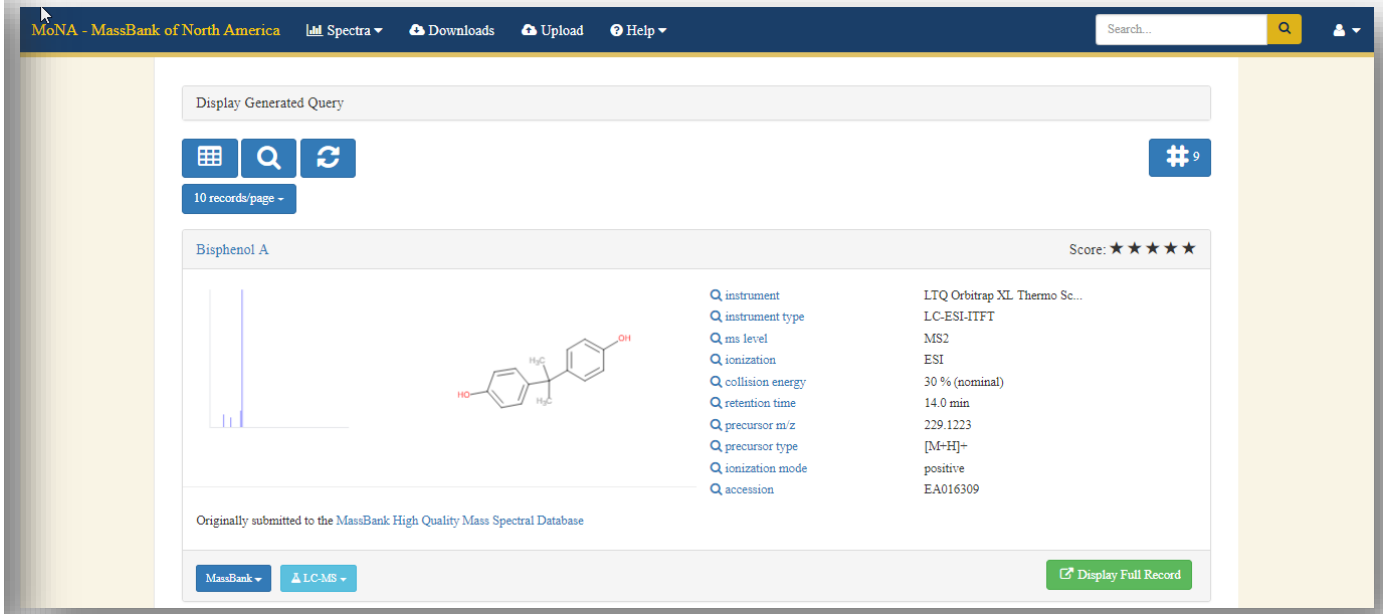
EPA United States Environmental Protection Agency



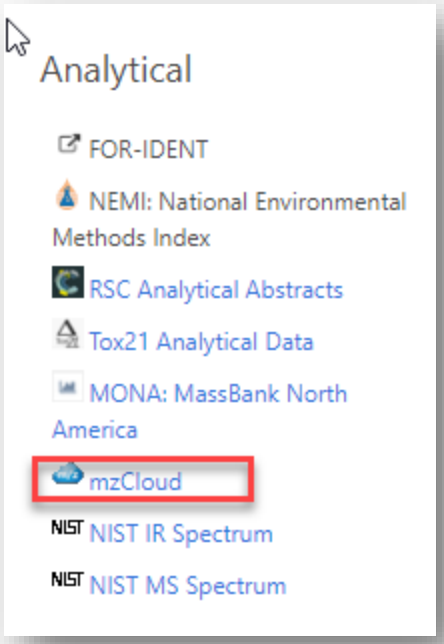


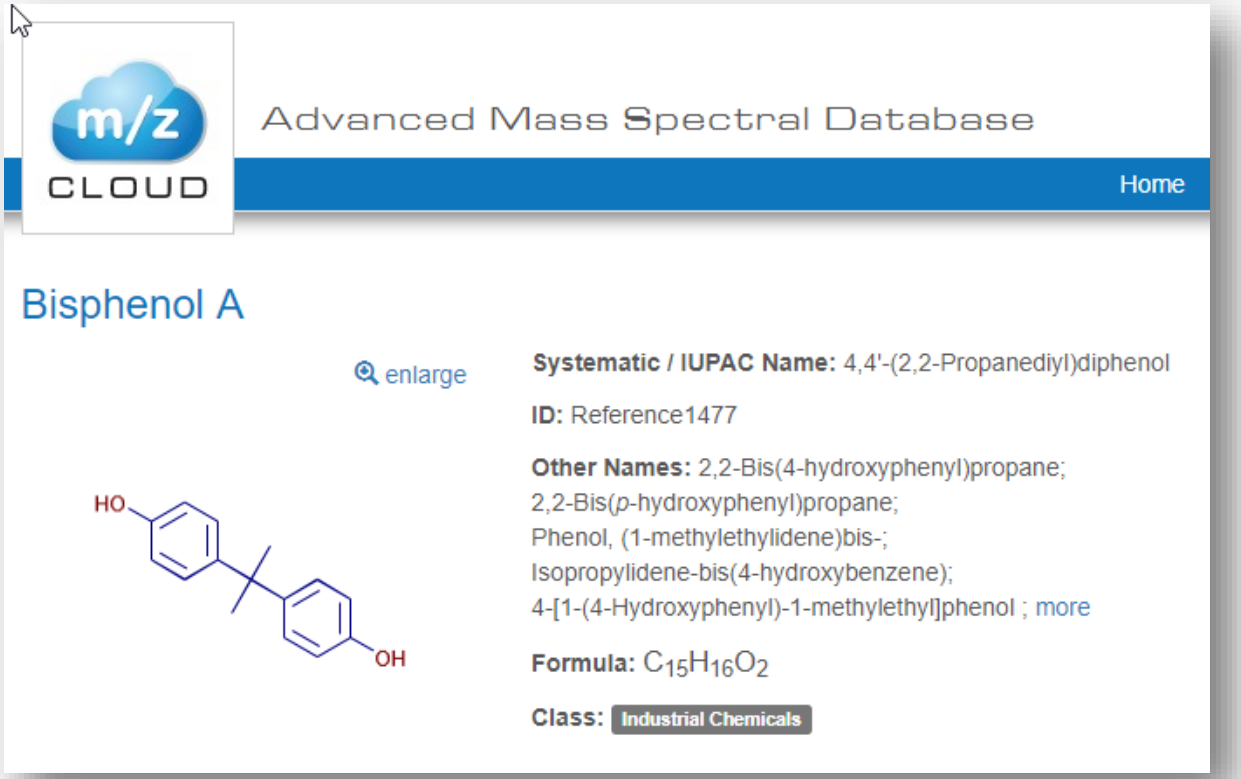
MassBank of North America
https://mona.fiehnlab.ucdavis.edu





m/z CLOUD
https://www.mzcloud.org/





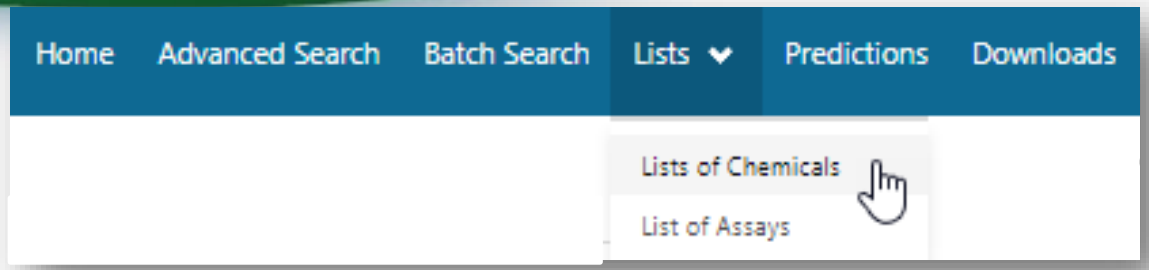
The broad dashboard is agnostic across chemicals

- This is fundamentally equivalent to sequential individual searches of other databases like NIST, ChemSpider, PubMed, etc.

- A main feature of the dashboard is a series of “Focused Chemical Lists”

FOCUSED CHEMICAL LISTS OF INTEREST

Chemical Lists



Download | Columns | mass | Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
HDXEXCH	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - Under HDX Conditions	2018-11-07	592	Observed species (deuterated and undeuterated) from the HDXNOEX list under hydrogen deuterium exchange conditions (Ruttkies, Schymanski et al. in prep.)
HDXNOEX	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - No Exchange	2018-11-07	765	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule high resolution mass spectrometry (Ruttkies, Schymanski et al. in prep.)
MASSBANKEUSP	MASSPECDB: MassBank.EU Collection: Special Cases	2017-07-16	263	The MassBank.EU list contains curated chemicals (Schymanski/Williams) associated with the literature/tentative/unknown/SI spectra available on MassBank.EU that are not available as part of the full MassBank collection of reference standard spectra.
MASSBANKREF	MASSPECDB: MassBank Reference Spectra Collection	2017-07-13	1267	This MassBank list contains chemicals associated with the full MassBank collection of reference standard spectra available on MassBank.EU, MassBank.JP and MassBank of North America as well as the Open Data collection, curated by Williams/Schymanski.
MYCOTOXINS	MASSPECDB: Mycotoxins from MassBank.EU	2017-08-02	88	This is a set of mycotoxins, initiated by the contribution of spectra of 90 mycotoxins to MassBank.EU by Justin Renaud and colleagues from Agriculture and Agri-Food Canada, Government of Canada

The Volatilome (ongoing growth)

- “Volatilome” compounds identified in breath (mostly gas-phase)
- Currently based on the deLacy Costello article in JBR 2014, and additions from EPA volatiles breath research – published and unpublished.

- de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, Osborne D, Ratcliffe NM, (2014). A review of the volatiles from the healthy human body. Journal of breath research. 2014 Jan 13;8(1):014001.
- Wallace MAG, Pleil JD, Oliver KD, Whitaker DA, Mentese S, Fent KW, and Horn GP, 2019. Non-targeted GC-MS analysis of exhaled breath samples: Exploring human biomarkers of exogenous exposure and endogenous response from professional firefighting activity. Journal of Toxicology and Environmental Health, Part A 82(4): 244-260.
- Pleil JD, Smith LB, and Zelnick SD, 2000. "Personal exposure to JP-8 jet fuel and exhaust at Air Force bases", Environmental Health Perspectives, 108(3): 183-192.
- Lindstrom AB, Pleil JD and Berkoff DC, 1997. "Alveolar breath sampling and analysis to assess trihalomethane exposures during competitive swimming training", Environmental Health Perspectives 105:6 636-642.

Volatilome Lists:

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
VOLATILOME	LIST: Subset of compounds detected in Human Breath	2019-07-11	777	A subset of compounds detected in human breath
VOLATILOME2	LIST: polar, semi-volatile, and condensed phase organic compounds found in human blood, condensed breath and urine	2019-07-17	133	This list is a subset of compounds detected in human biological media including blood, dried blood spots (DBS), urine, exhaled breath condensate (EBC), and exhaled breath aerosols (EBA).

Example from Volatilome list:

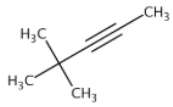
EPA United States Environmental Protection Agency

List Details

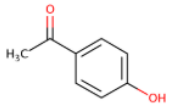
Description: This list is a subset of compounds detected in human breath and reported in the peer-reviewed literature and identified in experimental work at US-EPA. The bulk of the collection is extracted from the article "The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva" by de Lacy Costello et al in J. Breath Res. 8 (2014) 034001 (DOI:10.1088/1752-7155/8/3/034001) as well as an increasing number of chemicals identified in our own laboratory studies.
Number of Chemicals: 777

777 chemicals

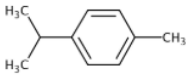
Select all | Download | Send to Batch Search | CASRN | CASRN | DTXSID | Mono.Mass | Hide chemicals that are: | Filter by Name or CASRN



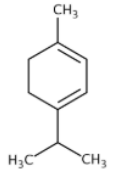
2-Pentyne, 4,4-dimethyl-
CASRN: 999-78-0
DTXSID: DTXSID40244289
Mono.Mass: 96.0939



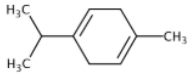
4-Hydroxyacetophenone
CASRN: 99-93-4
DTXSID: DTXSID0029133
Mono.Mass: 136.052429



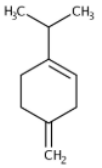
p-Cymene
CASRN: 99-87-6
DTXSID: DTXSID3026645
Mono.Mass: 134.10955



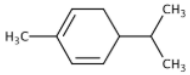
alpha-Terpinene
CASRN: 99-86-5
DTXSID: DTXSID9041237
Mono.Mass: 136.125201



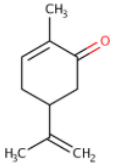
gamma-Terpinene
CASRN: 99-85-4
DTXSID: DTXSID6041210
Mono.Mass: 136.125201



b-Terpinene
CASRN: 99-84-3
DTXSID: DTXSID60243931
Mono.Mass: 136.125201



alpha-Phellandrene
CASRN: 99-83-2
DTXSID: DTXSID4047593
Mono.Mass: 136.125201



dl-Carvone
CASRN: 99-49-0
DTXSID: DTXSID8047426
Mono.Mass: 150.104465

Example for 86 Da integer mass:

LIST: Subset of compounds detected in Human Breath

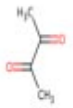
Search VOLATILOME Chemicals

Identifier substring search

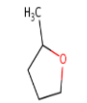
Description: This list is a subset of compounds detected in human breath and reported in the peer-reviewed literature and identified in experimental work at US-EPA. The bulk of the collection is extracted from the article "The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva" by de Lacy Costello et al in J. Breath Res. 8 (2014) 034001 (DOI:10.1088/1752-7155/8/3/034001) as well as an increasing number of chemicals identified in our own laboratory studies.
Number of Chemicals: 777

13 of 777 chemicals selected

Deselect all | Download | Send to Batch Search | Mono.Mass | Unselected | Filter by Name or CAS

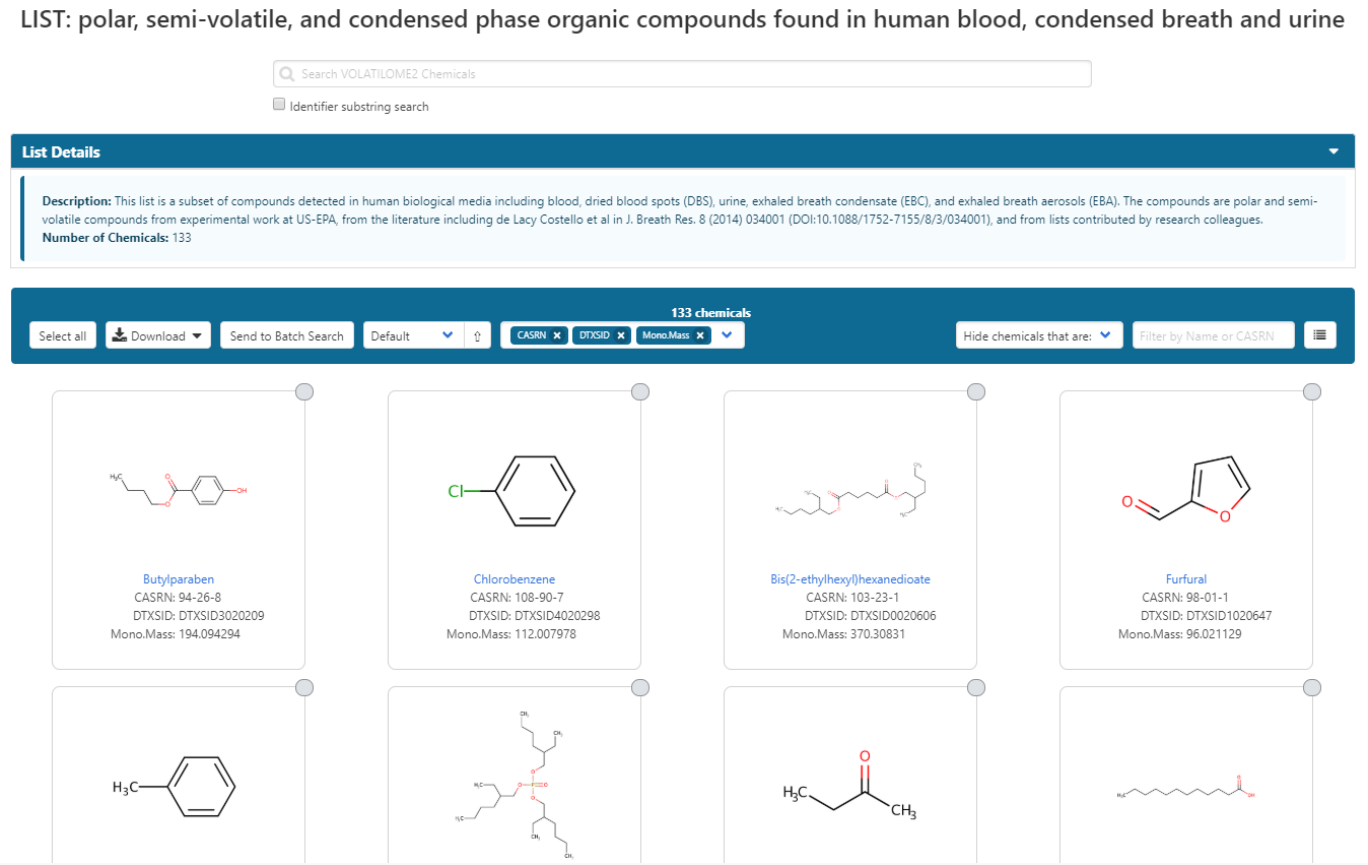
Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts	Monoisotopic Mass
	DTXSID6020224 ToxCast™	4-Butyrolactone	96-48-0	Level 1	191	101	125	3550	86.036779
	DTXSID6021583 ToxCast™	2,3-Butanedione	431-03-8	Level 2	25	86	111	1208	86.036779
	DTXSID1021619	3-Methylbutanal	590-86-3	Level 2	26	84	103	49	86.073165

Example for 86 Da integer mass:

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts	Monoisotopic Mass
	DTXSID7021653 ToxCast™	Pentanal	110-62-3	Level 2	25	101	103	43	86.073165
	DTXSID2021818 ToxCast™	2-Methylbutanal	96-17-3	Level 2	13	71	73	11	86.073165
	DTXSID0021888 ToxCast™	2-Pentanone	107-87-9	Level 2	30	98	111	19	86.073165
	DTXSID0022062 ToxCast™	3-Methyl-2-butanone	563-80-4	Level 2	13	86	104	6	86.073165
	DTXSID9030258 ToxCast™	2-Methyltetrahydrofuran	96-47-9	Level 2	4	62	109	0	86.073165
	DTXSID0021917	n-Hexane	110-54-3	Level 1	390	129	1027	1005	86.10955

Volatilome2 (ongoing growth)

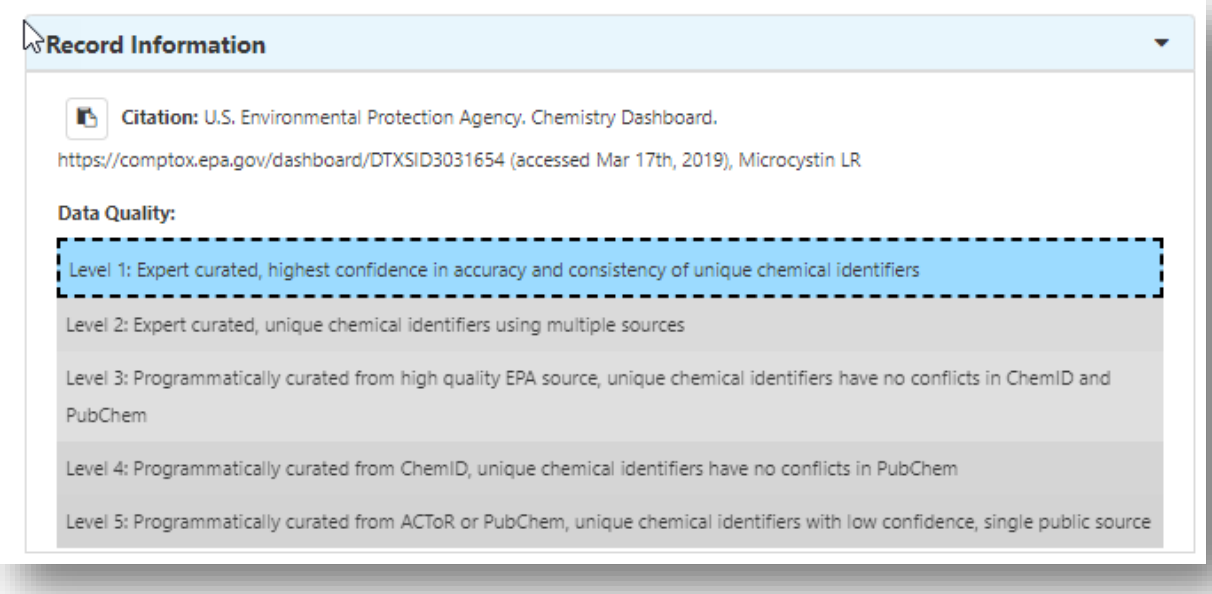
- “Volatilome2” - selection of compounds in human media including blood, urine, sweat, exhaled breath condensate and aerosol, as well as semi-volatile and non-volatile compounds.



DO WE REALLY NEED ANOTHER DATABASE?

Do we really need another database?

- The Dashboard is a tool to make research easier
- Many features are provided by various individual searches, but it would require a great deal of effort
- The data and chemical information on the Dashboard is under ongoing curation by EPA experts and assigned a level of confidence.



Is a bigger database better?

Journal of The American Society for Mass Spectrometry
January 2012, Volume 23, Issue 1, pp 179–185 | Cite as

Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider

Authors Authors and affiliations

James L. Little, Antony J. Williams, Alexey Pshenichnov, Valery Tkachenko

- ChemSpider was 26 million chemicals then
- Much BIGGER today
- Is bigger better??

71 Million chemical structures

Comparing Search Performance

Anal Bioanal Chem (2017) 409:1729–1735
DOI 10.1007/s00216-016-0139-z

RAPID COMMUNICATION

Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard

Andrew D. McEachran[1] · Jon R. Sobus[2] · Antony J. Williams[3]

- Dashboard content *was* 720k chemicals
- Only 3% of ChemSpider size
- What was the comparison in performance?

How did performance compare?

Summary statistics and rank-ordered position in the CompTox Chemistry Dashboard and ChemSpider of the 89 compound subset from the Little et al. [7] study

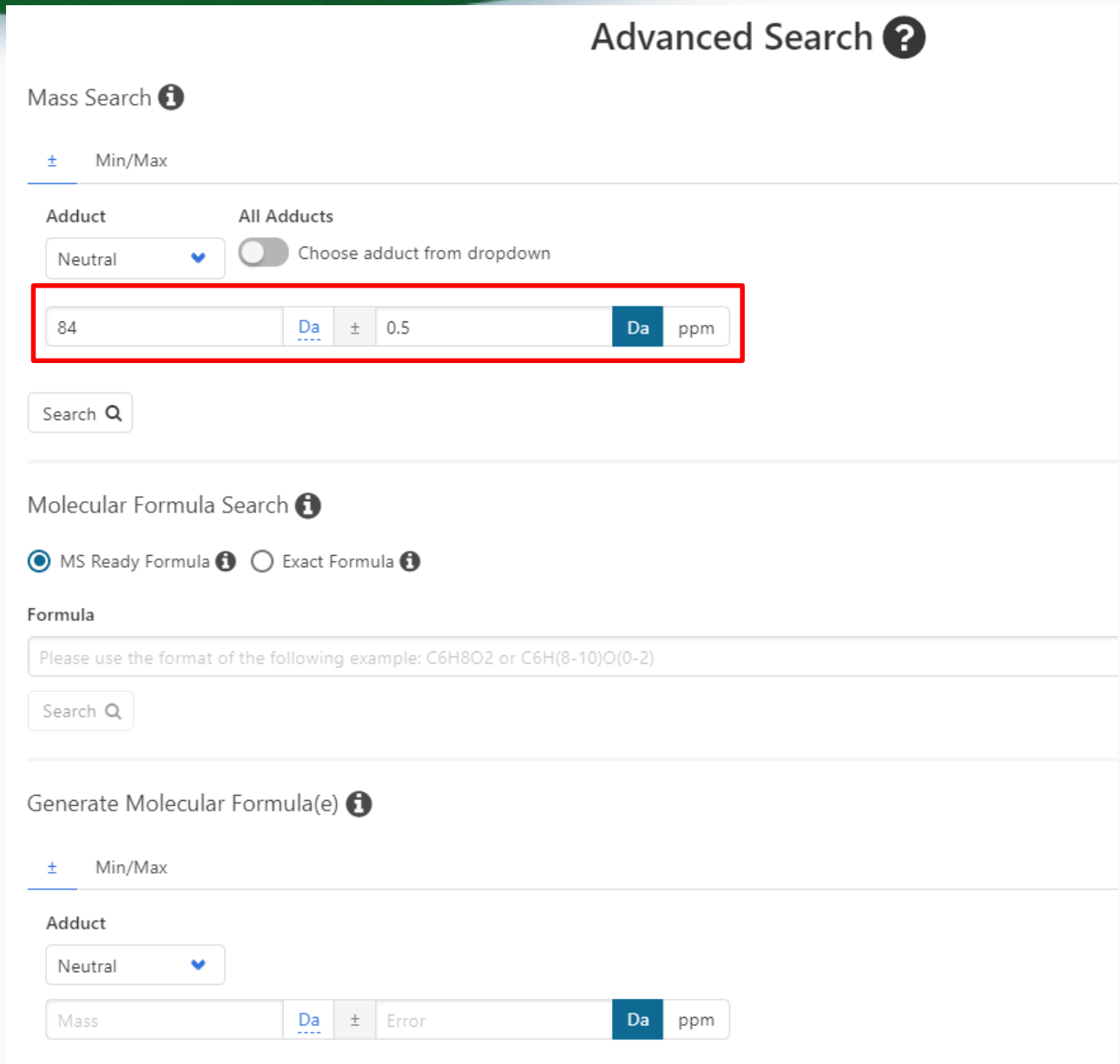
		Average rank	Number in each position rank-ordered				
		(±SD)	#1	#2	#3	#4	#5+
Mass-based	Dashboard	1.2 ± 0.7	77[a]	5	3	3	
	ChemSpider	2.2 ± 6.1[b]	68	8	7	1	5
Formula-based	Dashboard	1.1 ± 0.4	78[a]	8	2		
	ChemSpider	1.3 ± 1.0	77	8	2	1	2

[a]One chemical (tephrosin) not present in the Dashboard

MASS AND FORMULA SEARCHING

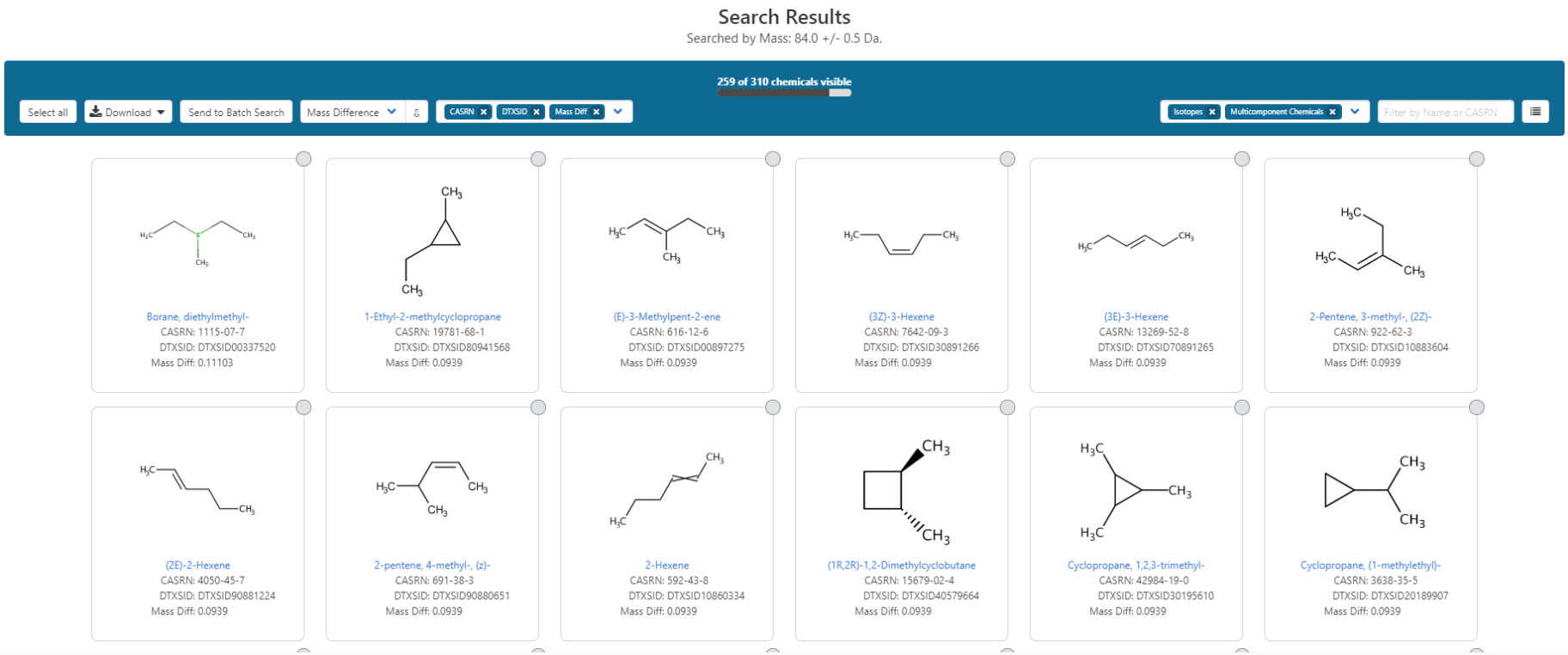
(and metadata ranking)

Advanced Searches
Mass and Formula Based Search



Advanced Searches
Mass and Formula Based Search

- Search 84+/-0.5 Da



Using Metadata for Ranking

- Use available metadata to rank candidates
  - Associated data sources
    - Associated lists in DSSTox database
    - Associated sources in PubChem
    - Specific types (e.g. water, surfactants, pesticides etc.)
  - Number of associated PubMed articles
  - Number of products/categories for the chemical

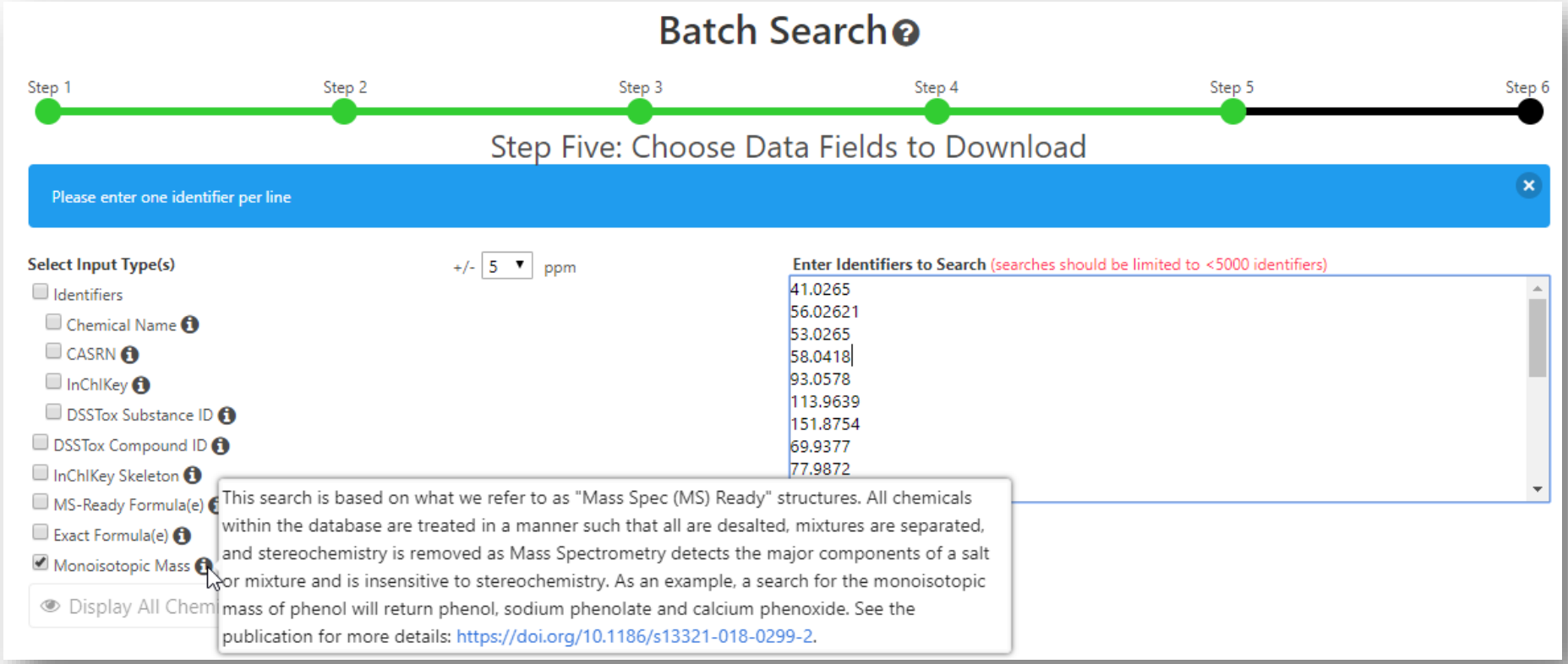
CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts	Monoisotopic Mass	Mass Difference
230	125	390	608	84.0939	0.0939
4	81	304	224	84.003371	0.003371

Batch Searching

- Singleton searches are useful but we work with **thousands** of masses and formulae!

- Typical questions
  - What is the list of chemicals for the formula $C_xH_yO_z$
  - What is the list of chemicals for a mass +/- error
  - Can I get chemical lists in Excel files? In SDF files?
  - Can I include properties in the download file?

Batch Searching Formula/Mass



Searching batches using MS-Ready Formula (or mass) searching

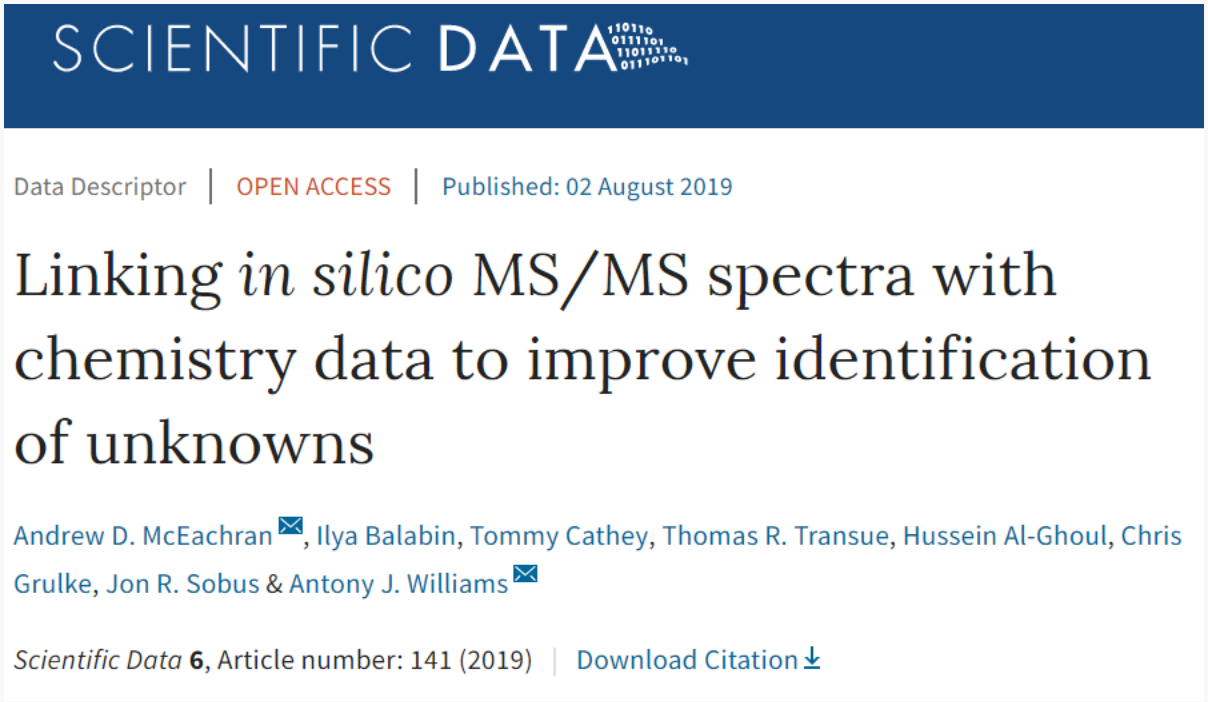
EPA United States Environmental Protection Agency

	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3	338.116398	19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3	266.163042576	19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine	C14H22N2O3	266.163042576	14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol	C14H22N2O3	266.163042576	12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide	C14H22N2O3	266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23ClN2O3	302.1397203	6
11	C14H22N2O3	DTXSID1068693	51963-82-7	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	C14H22N2O3	266.163042576	5
12	C18H34N2O6S	DTXSID3023215	154-21-2	Lincomycin	C18H34N2O6S	406.213757997	35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35ClN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM_71432748	C18H35ClN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O	176.094963014	40
16	C10H12N2O	DTXSID8075330	50-67-9	Serotonin	C10H12N2O	176.094963014	22
17	C10H12N2O	DTXSID8044412	2654-57-1	4-Methyl-1-phenylpyrazolidin-3-one	C10H12N2O	176.094963014	18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride	C10H13ClN2O	212.0716407	11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine	C10H12N2O	176.094963014	10
20	C10H12N2O	DTXSID10196105	443-31-2	6-Hydroxytryptamine	C10H12N2O	176.094963014	9
21	C10H12N2O	DTXSID90185693	31822-84-1	1,4,5,6-Tetrahydro-5-phenoxypyrimidine	C10H12N2O	176.094963014	7
22	C10H12N2O	DTXSID40178777	2403-66-9	2-Benzimidazolepropanol	C10H12N2O	176.094963014	7
23	C10H12N2O	DTXSID80157026	13140-86-8	N-Cyclopropyl-N'-phenylurea	C10H12N2O	176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9	4-Hydroxytryptamine	C10H12N2O	176.094963014	6
25	C14H18N4O3	DTXSID5023900	17804-35-2	Benomyl	C14H18N4O3	290.137890456	68
26	C14H18N4O3	DTXSID3023712	738-70-5	Trimethoprim	C14H18N4O3	290.137890456	51
27	C14H18N4O3	DTXSID40209671	60834-30-2	Trimethoprim hydrochloride	C14H19ClN4O3	326.1145682	8
28	C14H18N4O3	DTXSID70204210	55687-49-5	Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2,	C14H18N4O3	290.137890456	5
29	C14H18N4O3	DTXSID20152671	120075-57-2	6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quin	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-	C14H18N4O3	290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxypheny	C19H27N5O7	437.191048229	3
34	C14H18N4O3	DTXSID20143781	101204-93-7	1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam	C14H18N4O3	290.137890456	3
35	C12H11N7	DTXSID6021373	396-01-0	Triamterene	C12H11N7	253.107593382	52
36	C12H11N7	DTXSID00204465	5587-93-9	Ampyrimine	C12H11N7	253.107593382	7
37	C12H11N7	DTXSID5064621	7300-26-7	Benzenamine, 4-azido-N-(4-azidophenyl)-	C12H9N7	251.091943318	4
38	C12H11N7	DTXSID00848025	90293-82-6	Sulfuric acid--6-phenylpteridine-2,4,7-triamine (1/1)	C12H13N7O4S	351.074973101	1
39	C12H11N7	DTXSID50575293	92310-83-3	(1E)-N-Phenyl-1,2-bis(1H-1,2,4-triazol-1-yl)ethan-1-imine	C12H11N7	253.107593382	1
40	C8H9NO2	DTXSID2020006	103-90-2	Acetaminophen	C8H9NO2	151.063328534	75
41	C8H9NO2	DTXSID6026567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	60

WORK IN PROGRESS

Work in Progress

- Predicted Spectra for candidate ranking
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database



SCIENTIFIC DATA

Data Descriptor | OPEN ACCESS | Published: 02 August 2019

Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

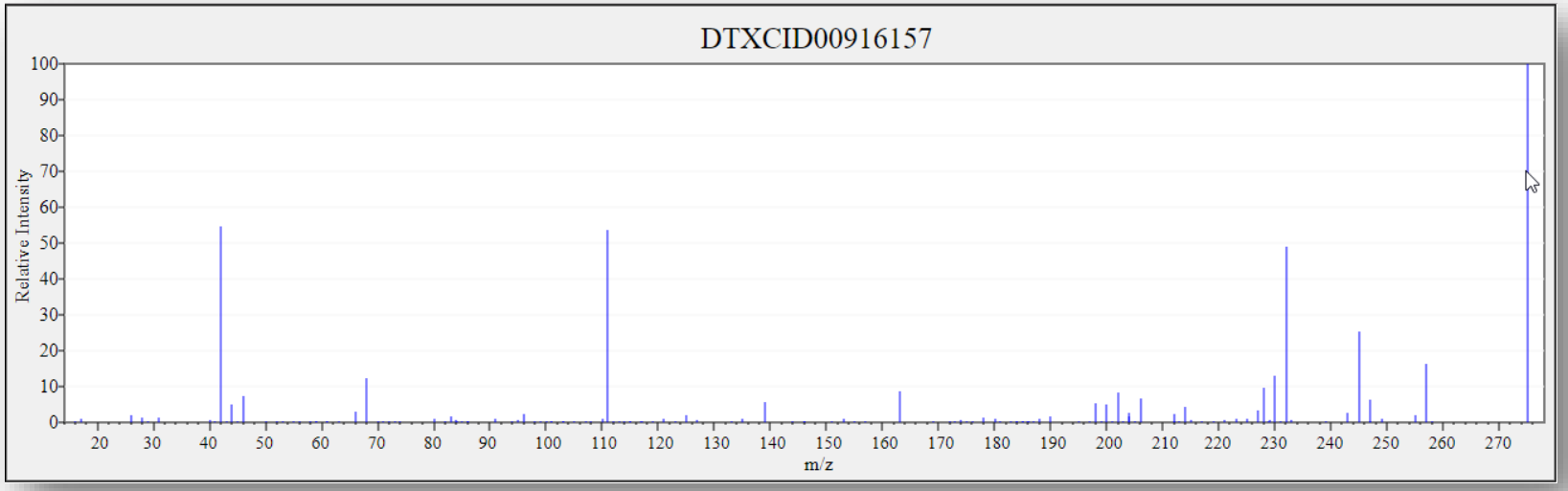
Andrew D. McEachran, Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghoul, Chris Grulke, Jon R. Sobus & Antony J. Williams

Scientific Data **6**, Article number: 141 (2019) | Download Citation

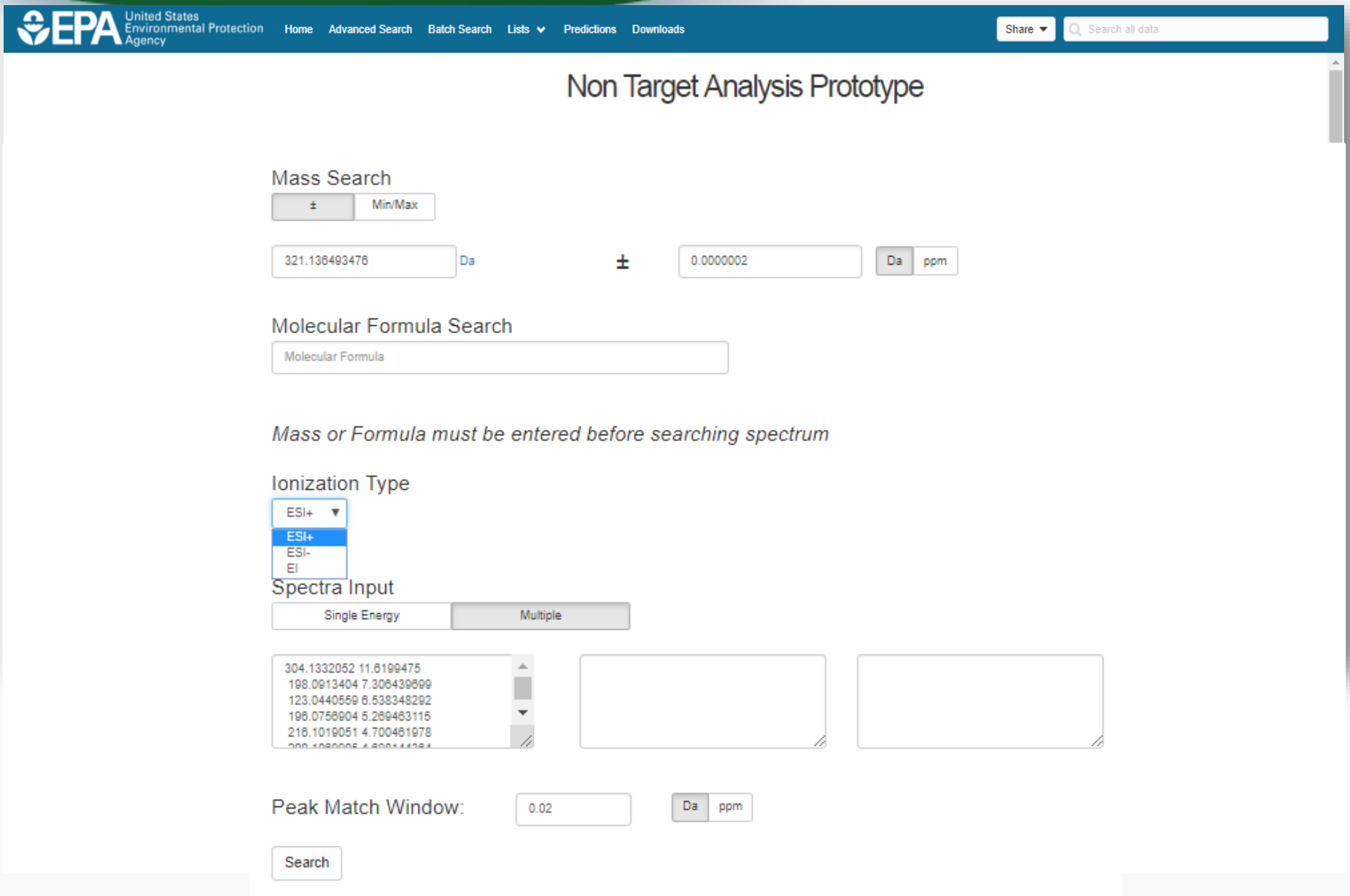
Predicted Mass Spectra

http://cfmid.wishartlab.com/

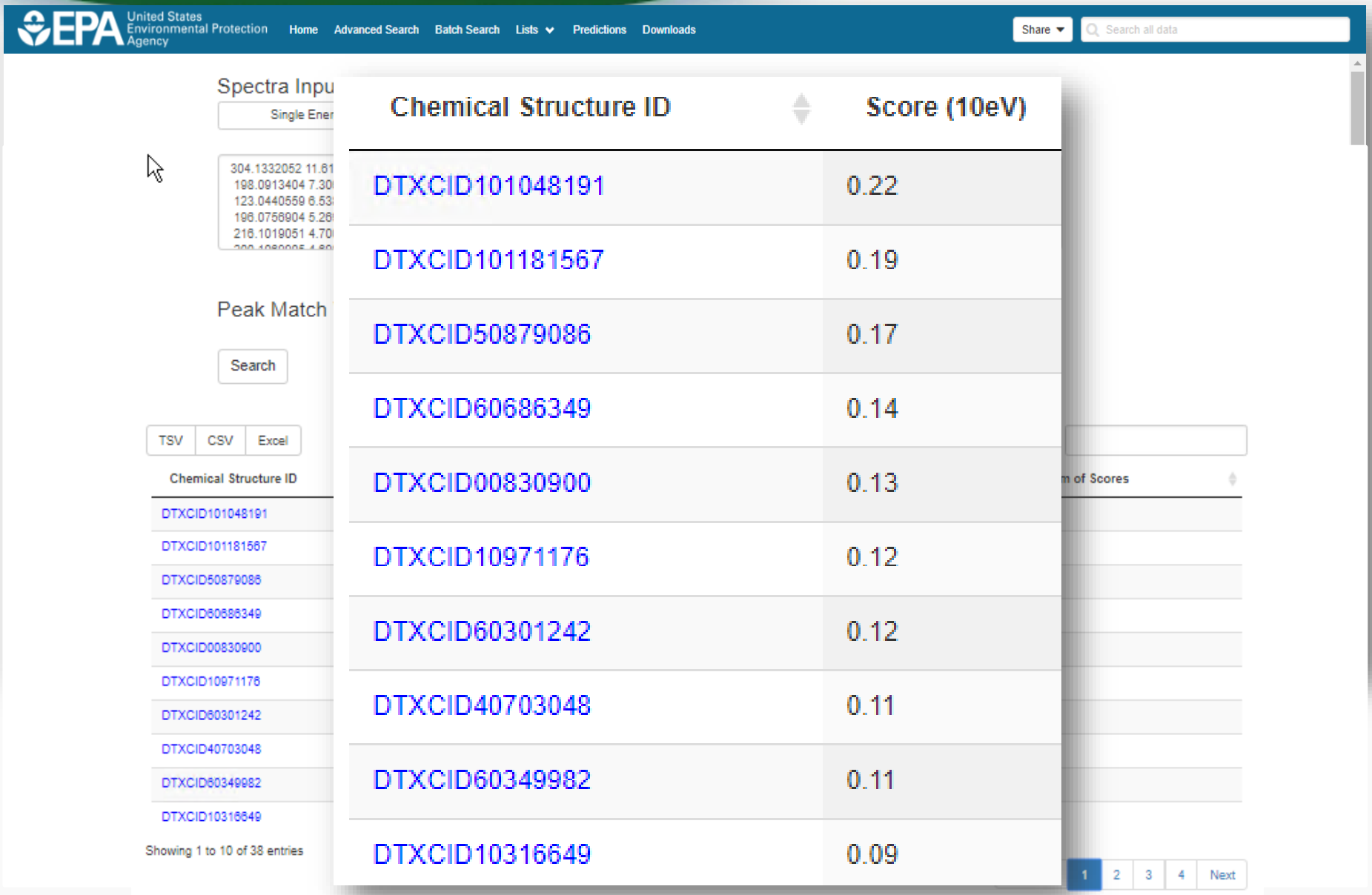
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra

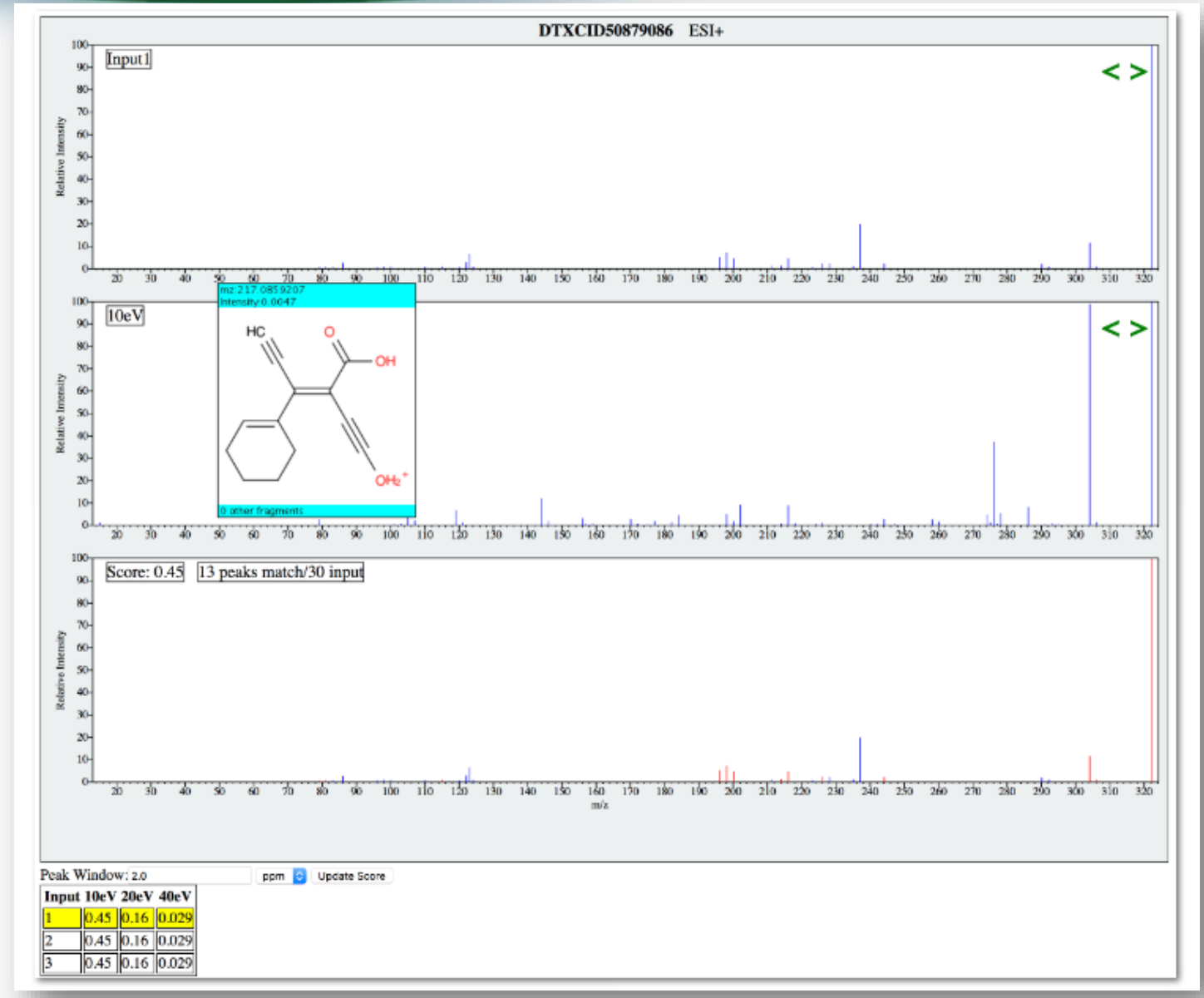


Search Expt. vs. Predicted Spectra



Chemical Structure ID	Score (10eV)
DTXCID101048191	0.22
DTXCID101181567	0.19
DTXCID50879086	0.17
DTXCID60686349	0.14
DTXCID00830900	0.13
DTXCID10971176	0.12
DTXCID60301242	0.12
DTXCID40703048	0.11
DTXCID60349982	0.11
DTXCID10316649	0.09

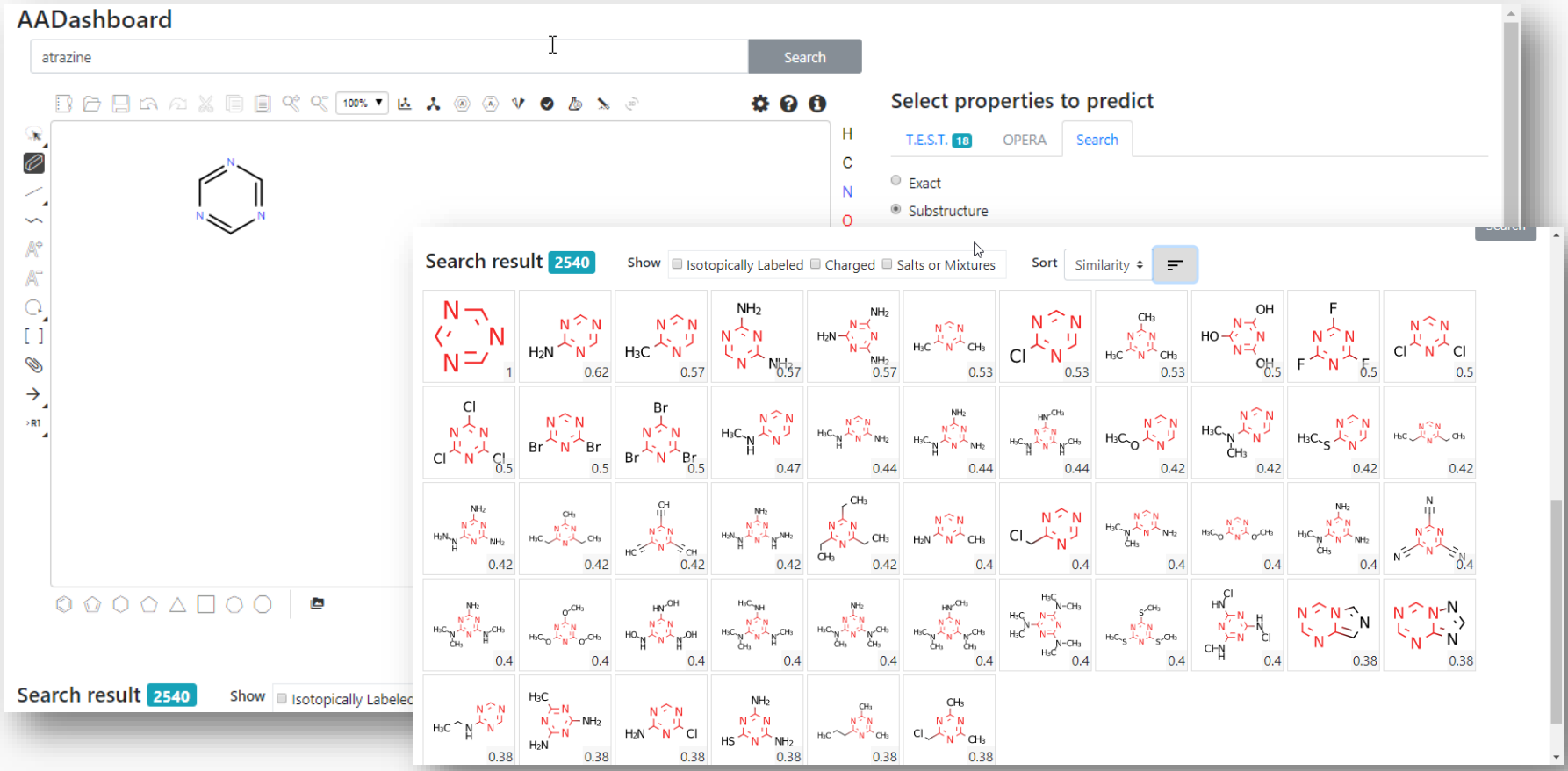
Spectral Viewer Comparison



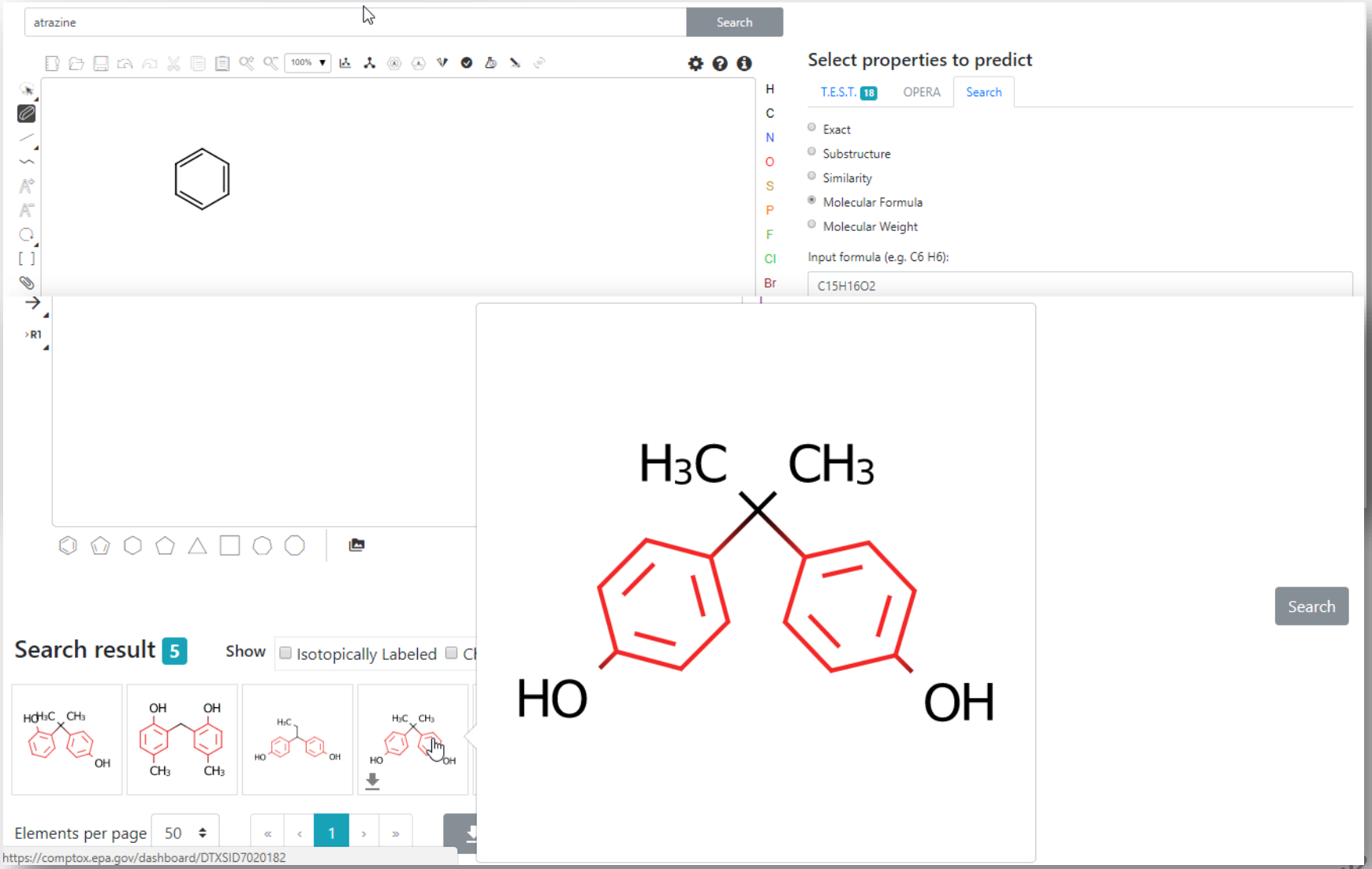
Work in Progress

- Predicted Spectra for candidate ranking
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Structure/substructure/similarity search

Prototype Development



Prototype Development

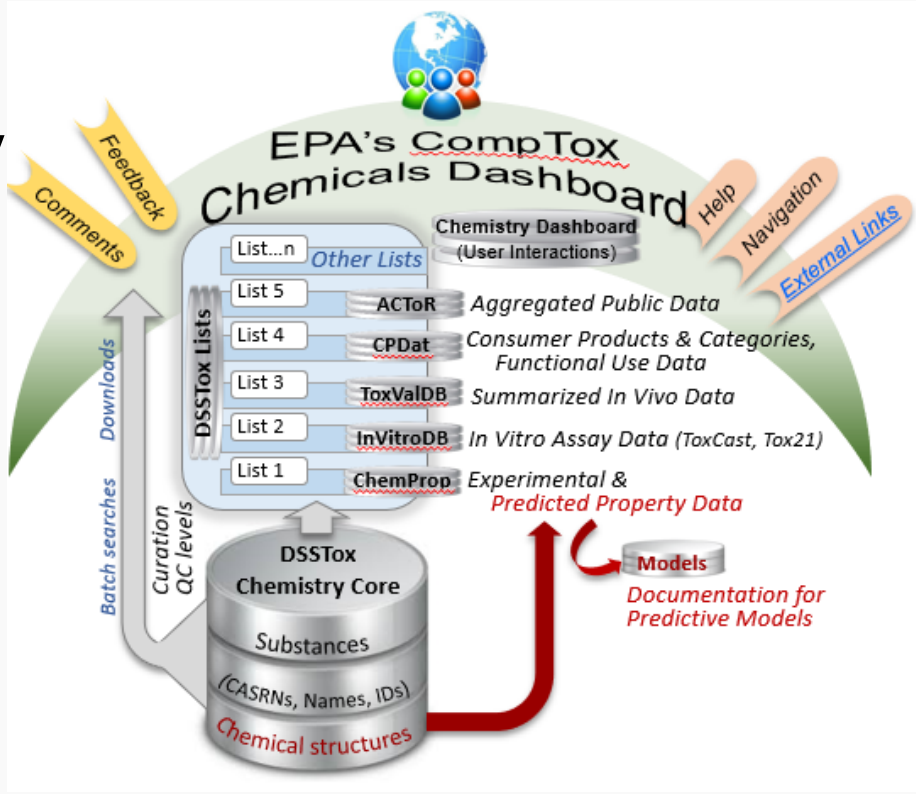


The Dashboard for Breath research

- We want a list of compounds found for exhaled breath condensate, or urine, or breath, etc.
- We want info about a single candidate chemical in an analysis
- We find a particular feature (base peak mass, molecular ion, spectrum) and we want to tentatively assign it (NTA)
- We find a series of “features” in analyses that may represent a variety of chemicals with similar mass, but possibly different formulas (suspect analysis).
- Link identifications by library searches to the likelihood that the chemical would be in the particular matrix and also measurable by the particular instrumentation.

Conclusion

- Dashboard access to data for ~875,000 chemicals
- MS-Ready data facilitates structure identification
- Related metadata facilitates candidate ranking
- Relationship mappings and chemical lists of great utility
- Dashboard and contents are one part of the solution



Solicitation for chemicals to include in future

- The lists in the dashboard are updated twice a year with newly curated chemicals.
- We are soliciting breath chemicals (gas-phase) for the Volatilome List.
- We are soliciting chemicals (condensate, aerosol) and other liquid based chemicals (blood, urine) for the Volatilome 2 list.
- We may further sub-divide lists if expedient.

Contact

Antony Williams
US EPA Office of Research and Development
National Center for Computational Toxicology
EMAIL: Williams.Antony@epa.gov

Joachim Pleil
US EPA Office of Research and Development
National Exposure Research Laboratory
EMAIL: Pleil.Joachim@epa.gov