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

CompTox chemicals dashboard: Data and tools to support chemical and environmental risk assessment and the ENTACT project

Antony Williams[1], *Chris Grulke*[1], *Andrew McEachran*[2], *Emma Schymanski*[3], *Jon Sobus*[4], *Elin Ulrich*[4], *Ann Richard*[1], *Jeremy Dunne*[1] *and Jeff Edwards*[1]

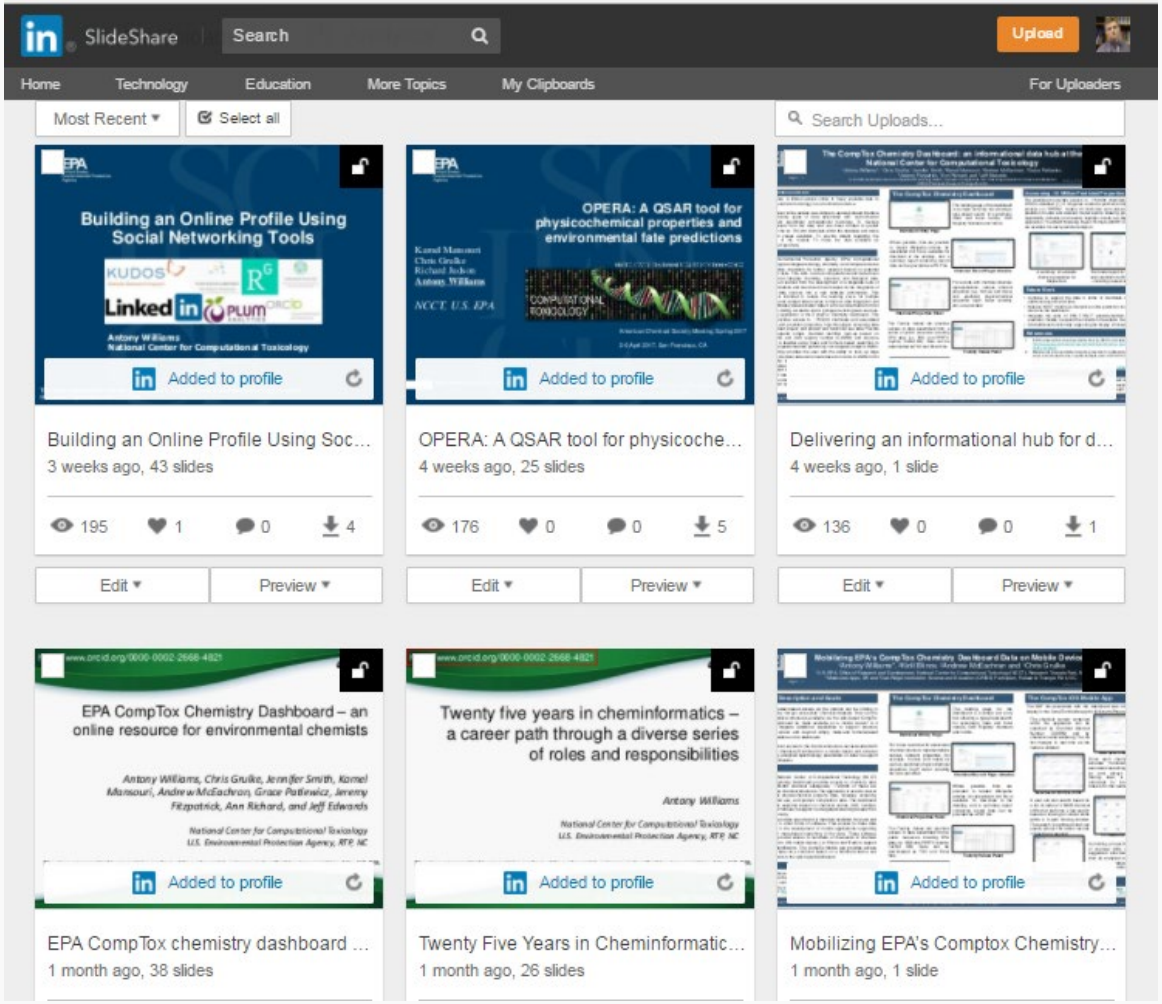
1) *National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC*
2) *Oak Ridge Institute of Science and Education (ORISE) Research Participant, RTP, NC*
3) *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg*
4) *National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC*

Spring 2019
ACS Spring Meeting, Orlando

Limit notetaking if you wish

EPA United States Environmental Protection Agency

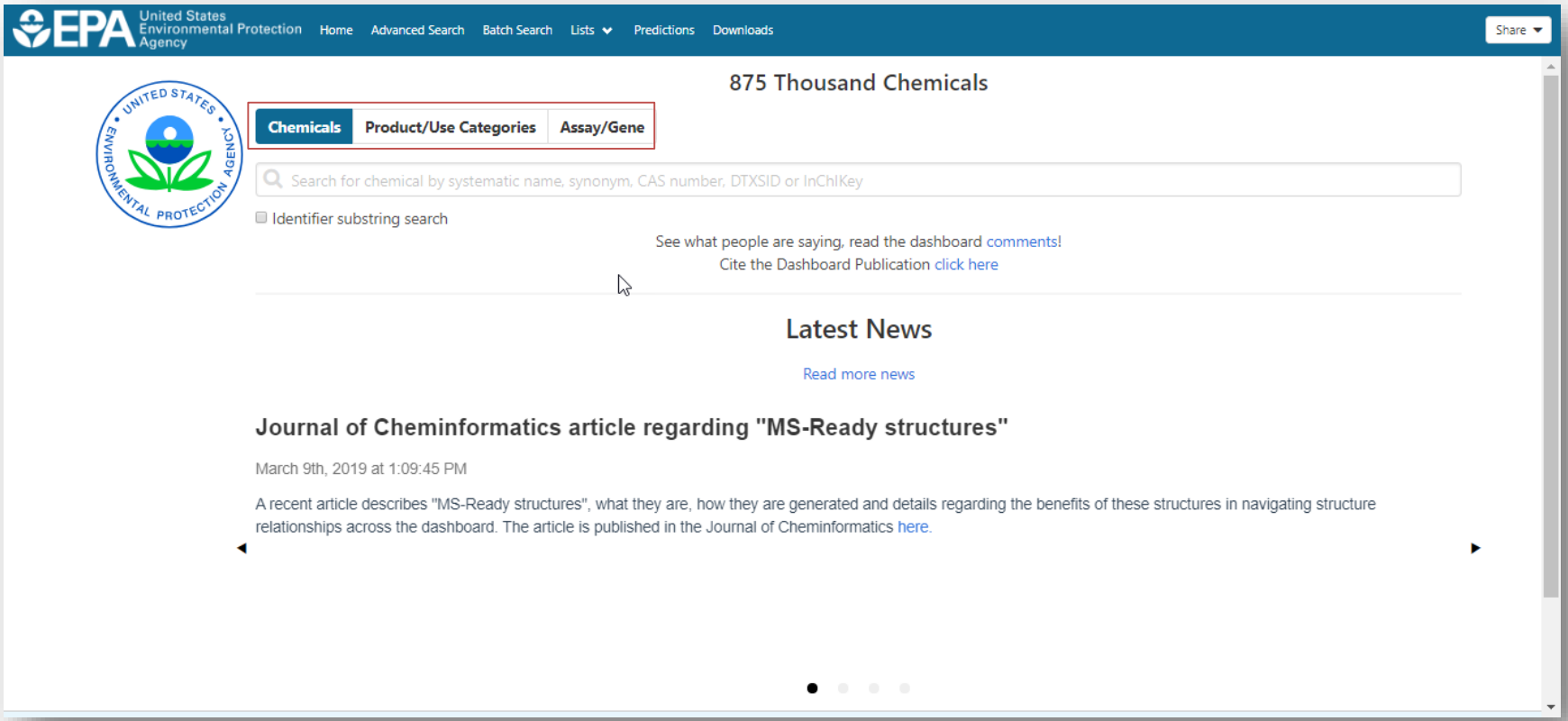
www.slideshare.net/AntonyWilliams



CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard

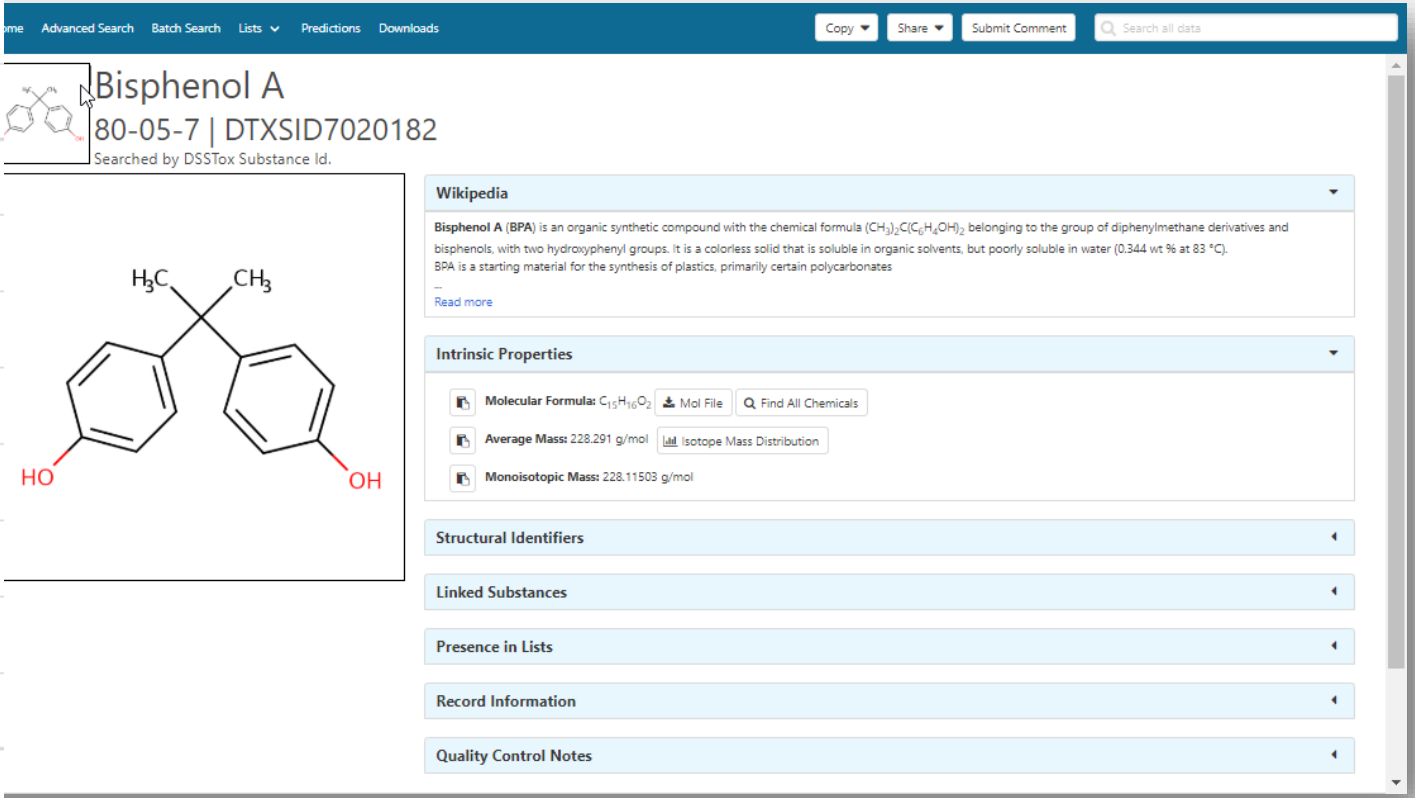
875k Chemical Substances



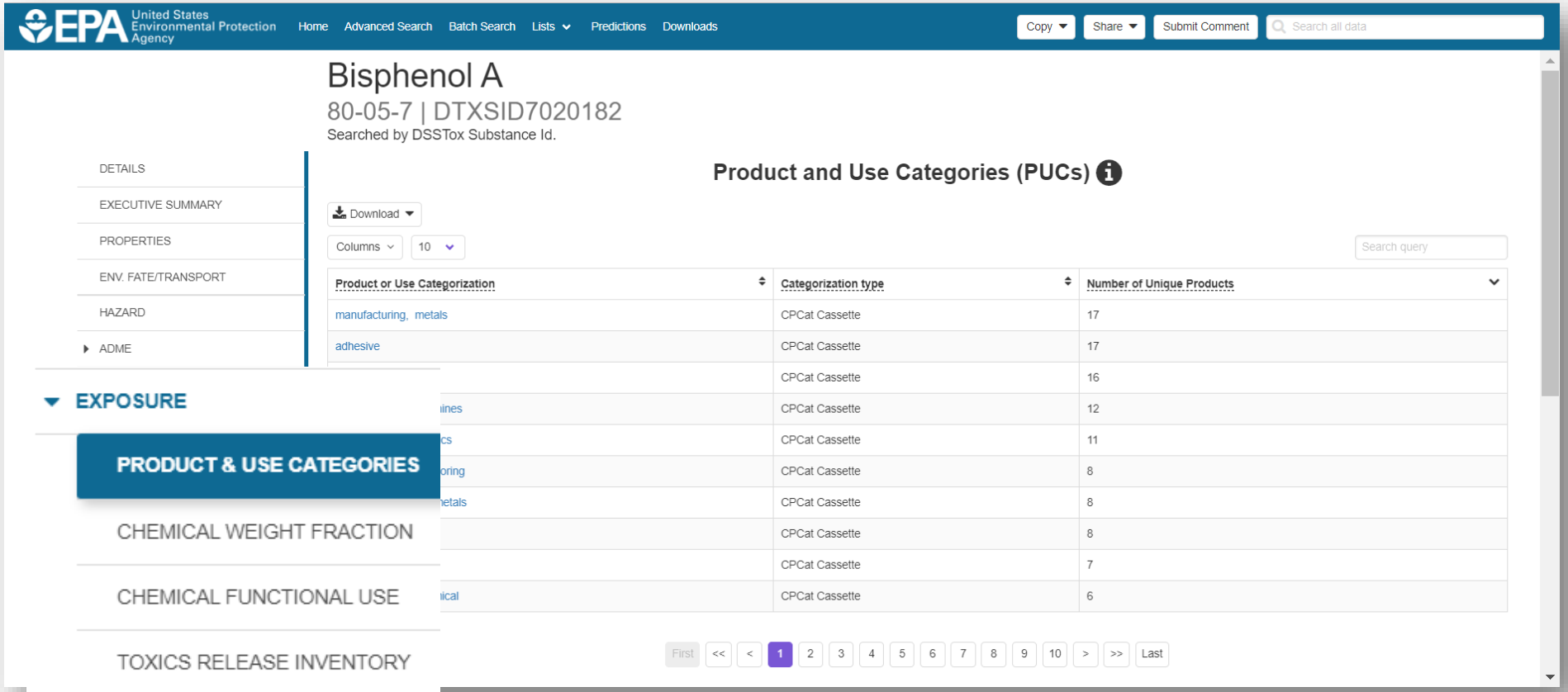
Detailed Chemical Pages

EPA United States Environmental Protection Agency

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



Sources of Exposure to Chemicals

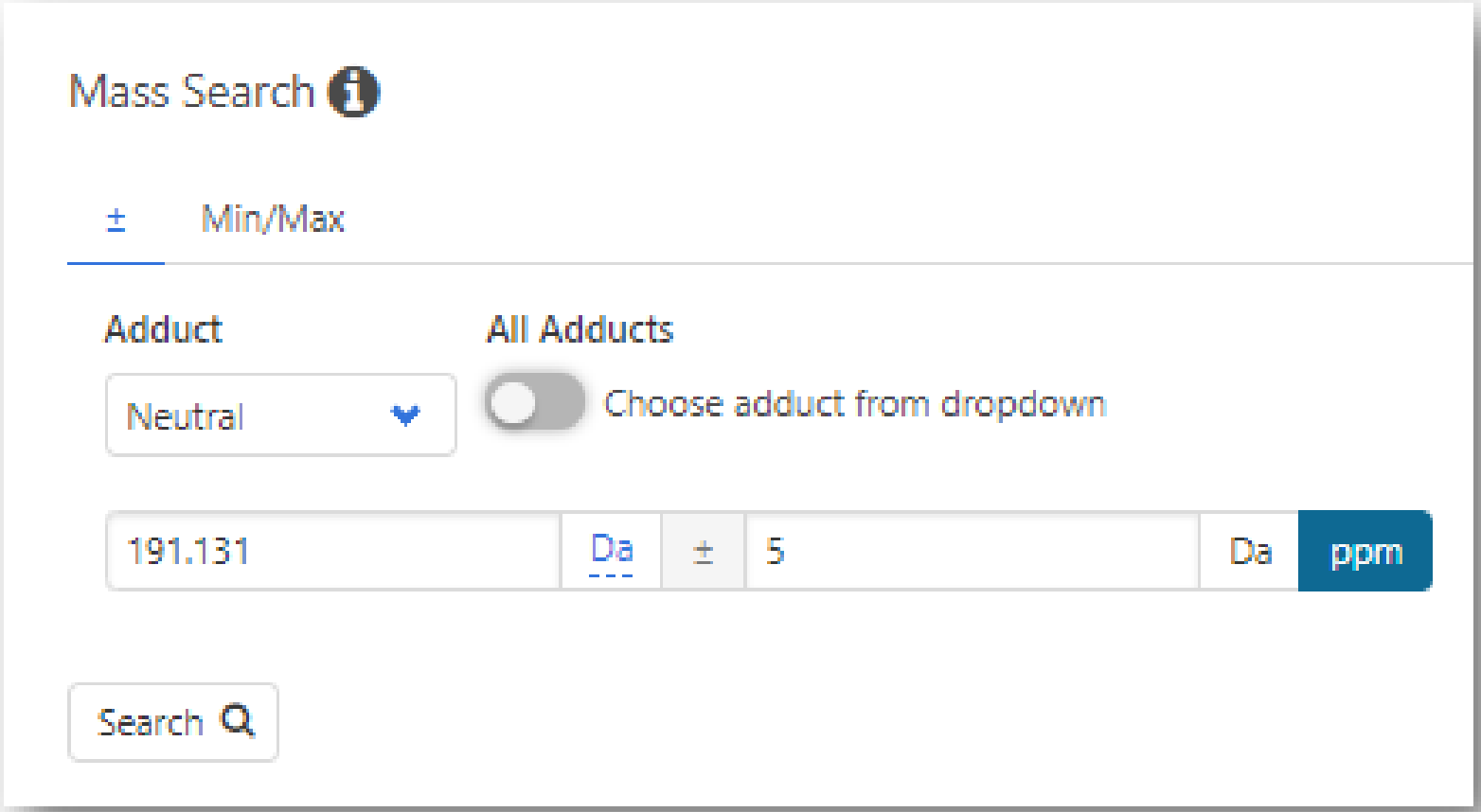


Mass Spec Specific Functionality

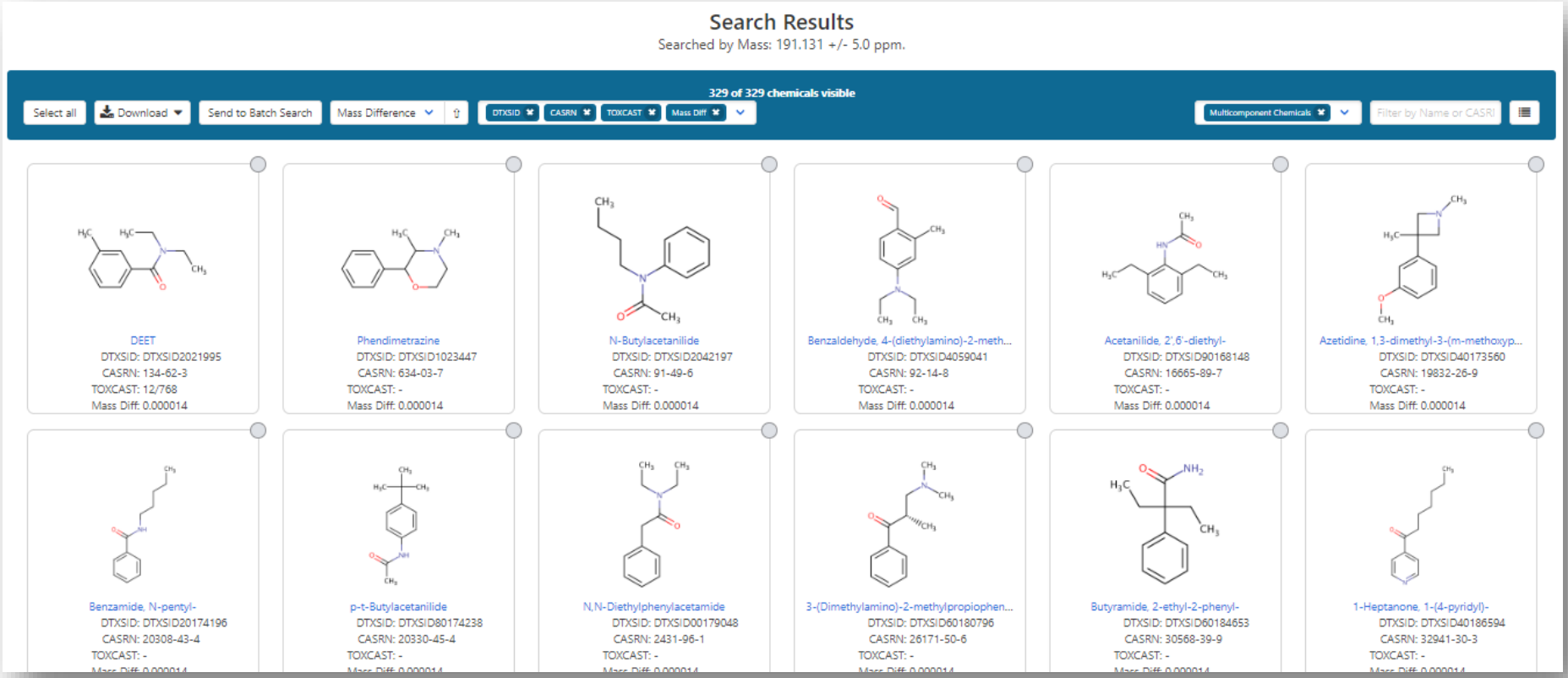
- The ENTACT project has driven specific functionality Mass Spectrometry needs
  - Specific searches within the dashboard
  - Mappings between chemicals in the database
  - Making use of specific representations for UVCBs
  - Addition of specific lists of chemicals
  - Collaborative cross-linking between sites
  - Research into best approaches for candidate ranking

Mass/Formula Searching and Metadata Ranking

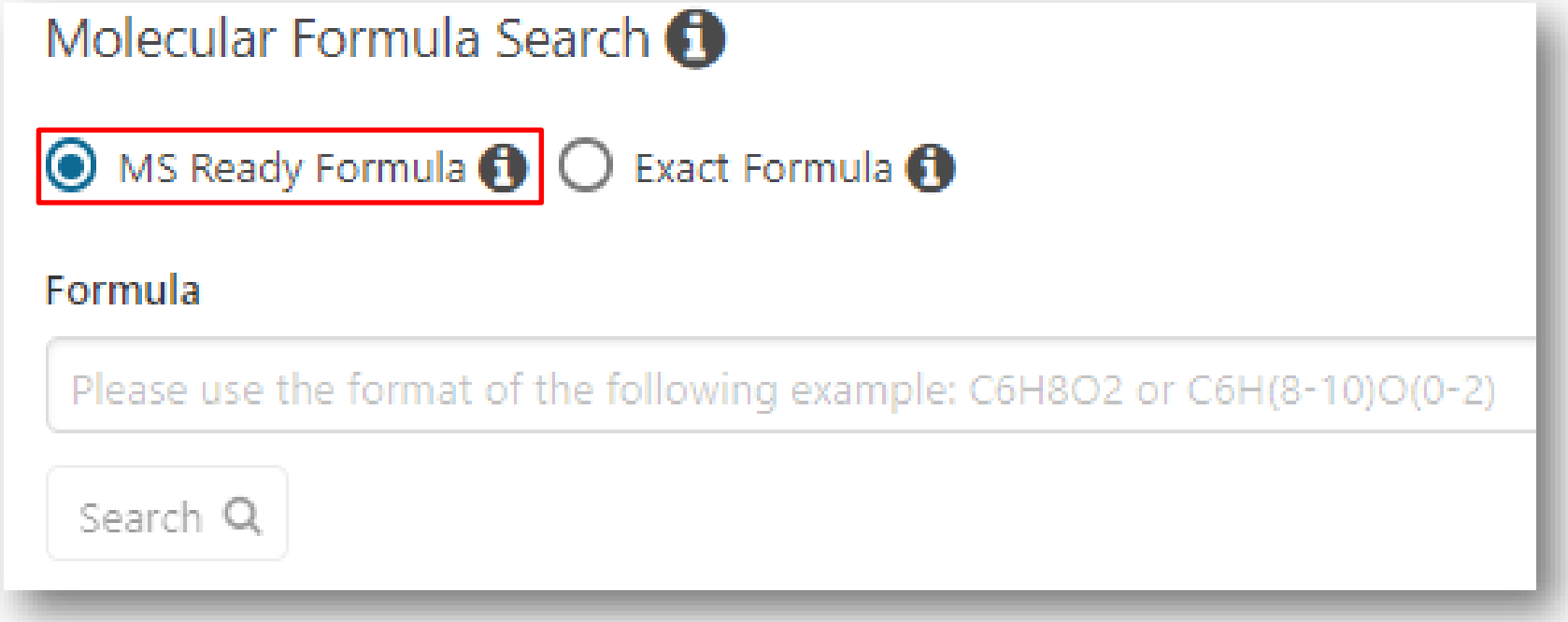
Advanced Searches
Mass Search



Advanced Searches
Mass Search



MS-Ready Structures for **Formula Search**



“MS-Ready Structures”

https://doi.org/10.1186/s13321-018-0299-2

McEachran *et al. J Cheminform (2018) 10:45*
https://doi.org/10.1186/s13321-018-0299-2

Journal of Cheminformatics

METHODOLOGY Open Access

“MS-Ready” structures for non-targeted high-resolution mass spectrometry screening studies

Andrew D. McEachran[1,2*], Kamel Mansouri[1,2,3], Chris Grulke[2], Emma L. Schymanski[4], Christoph Ruttkies[5] and Antony J. Williams[2*]

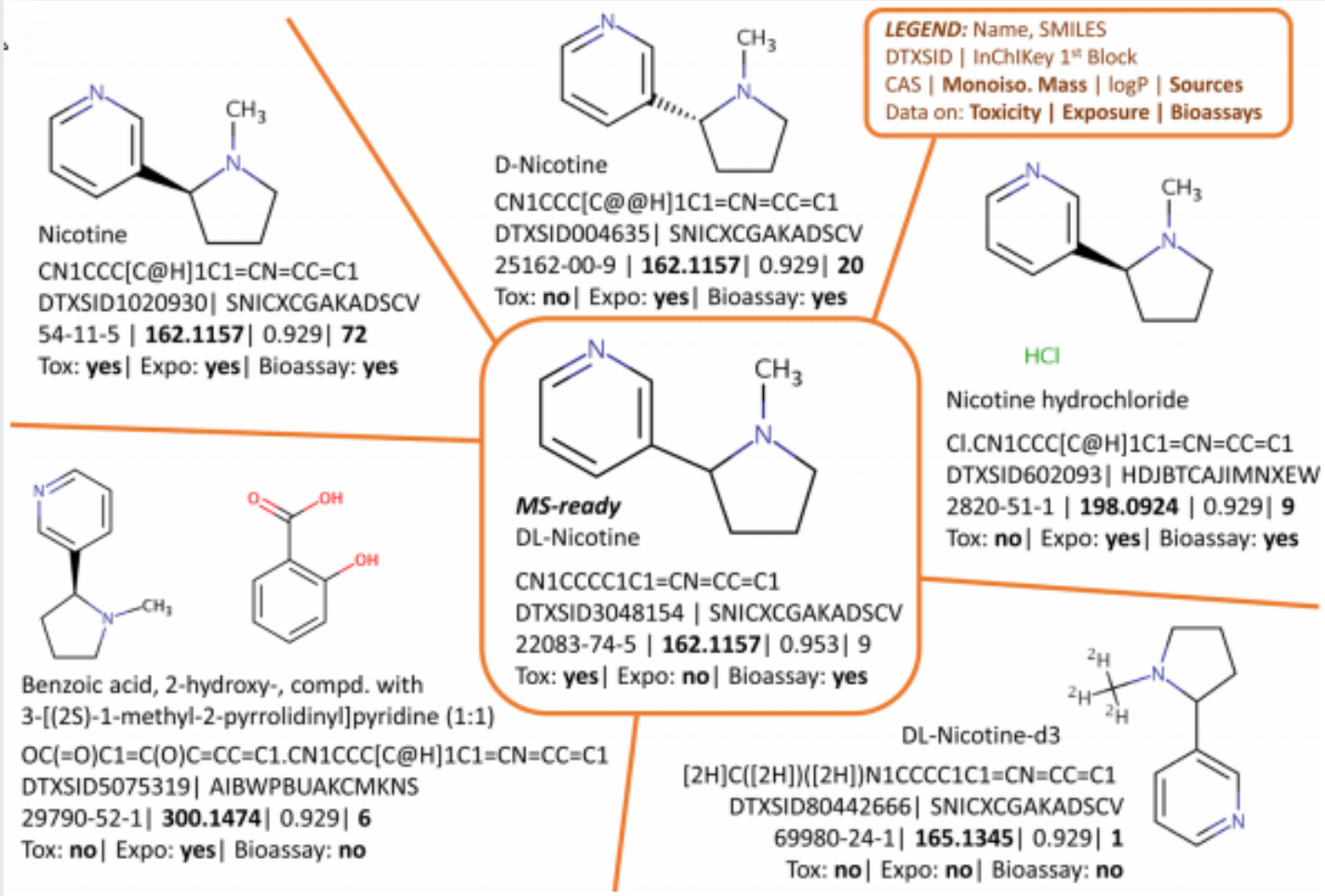
Environmental Science & Technology

Viewpoint

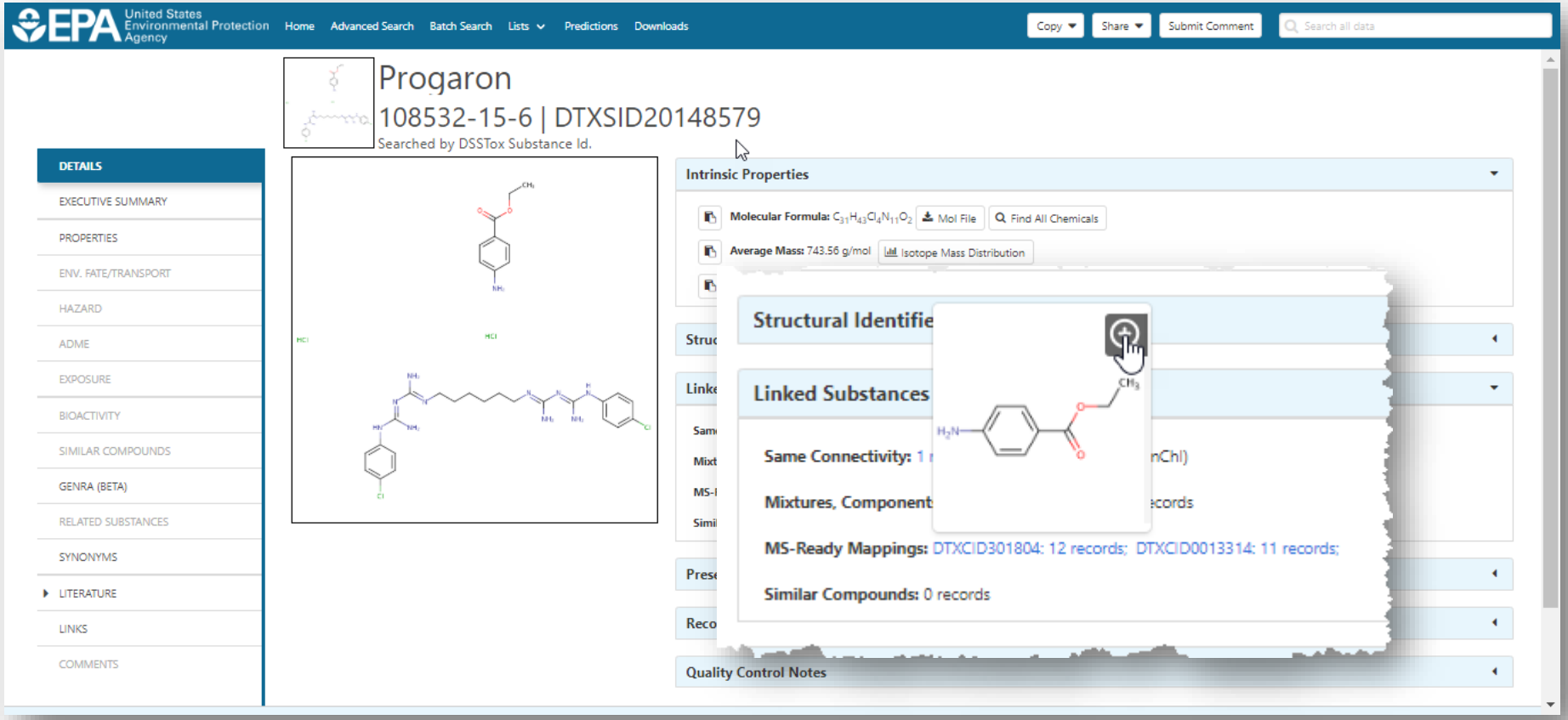
pubs.acs.org/est

Open Science for Identifying "Known Unknown" Chemicals

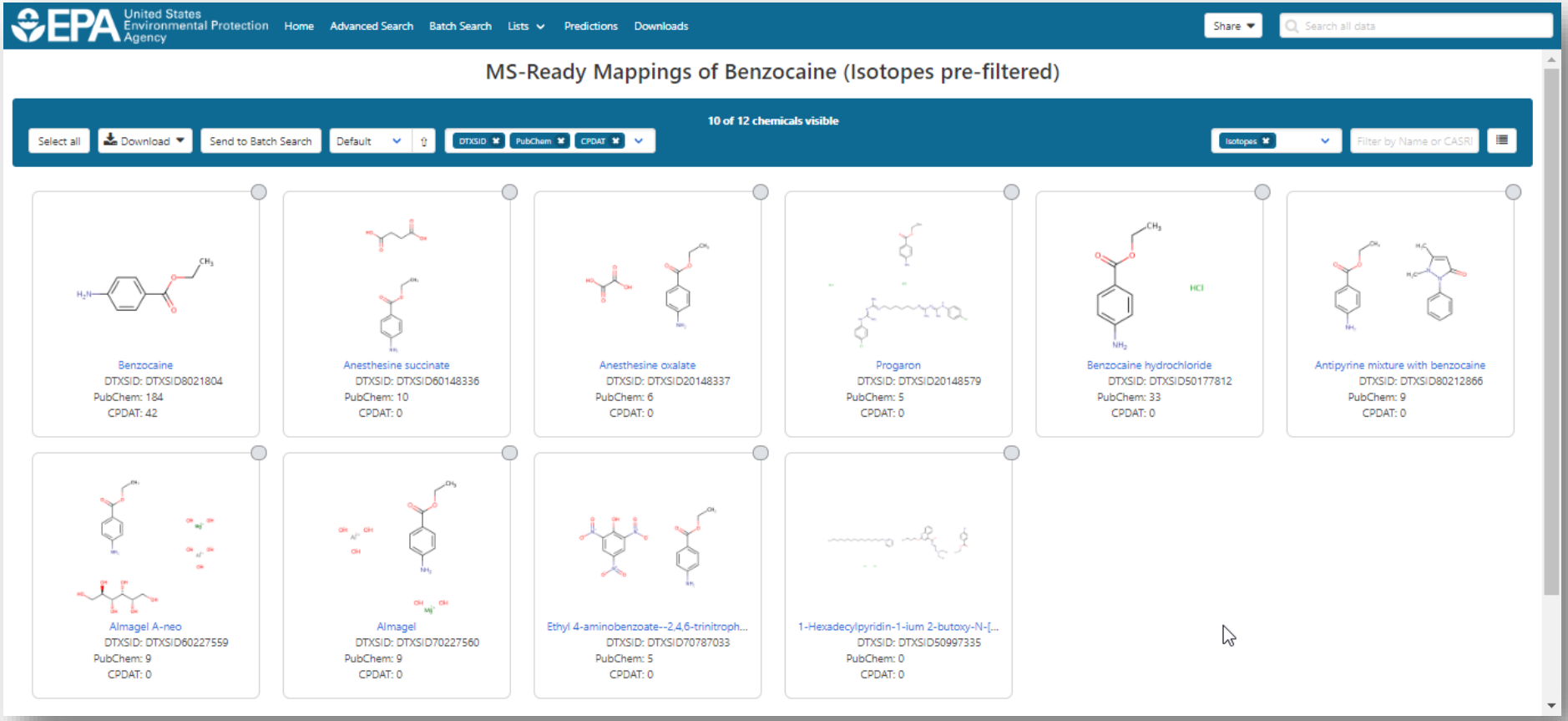
Emma L. Schymanski*,† and Antony J. Williams*,‡



MS-Ready Mappings

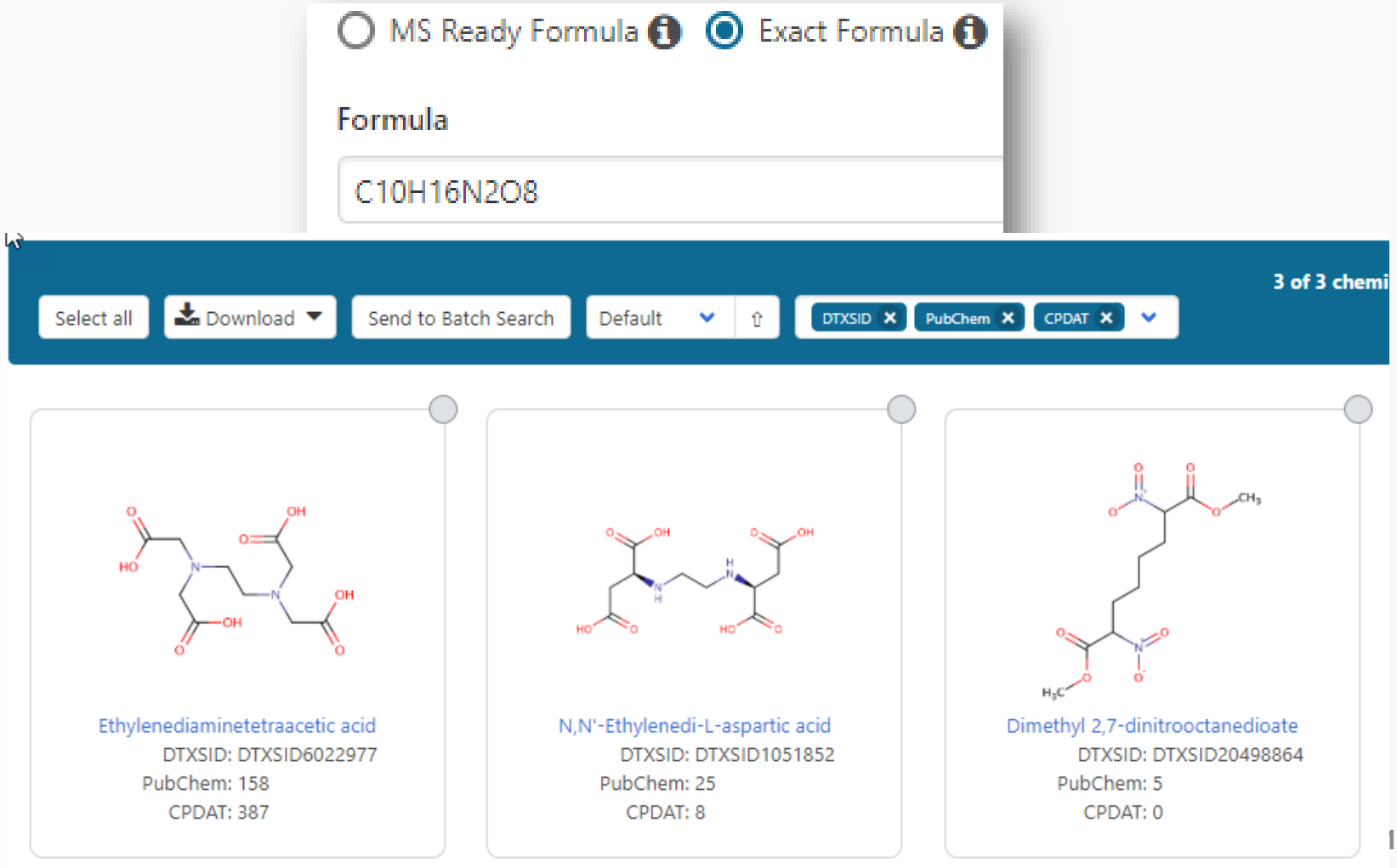


MS-Ready Mappings Set



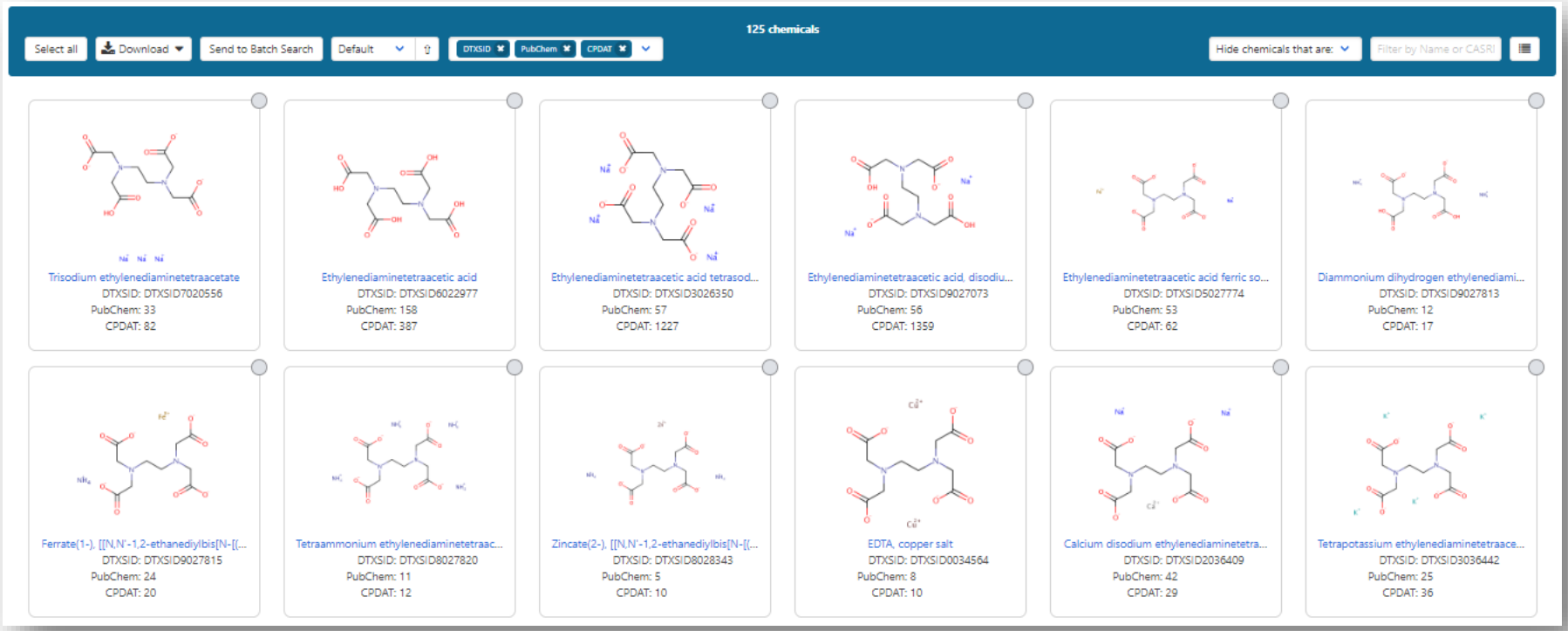
MS-Ready Mappings

- **EXACT Formula**: C10H16N2O8: 3 Hits



MS-Ready Mappings

- **Same** Input Formula: C10H16N2O8
- **MS Ready Formula** Search: **125** Chemicals

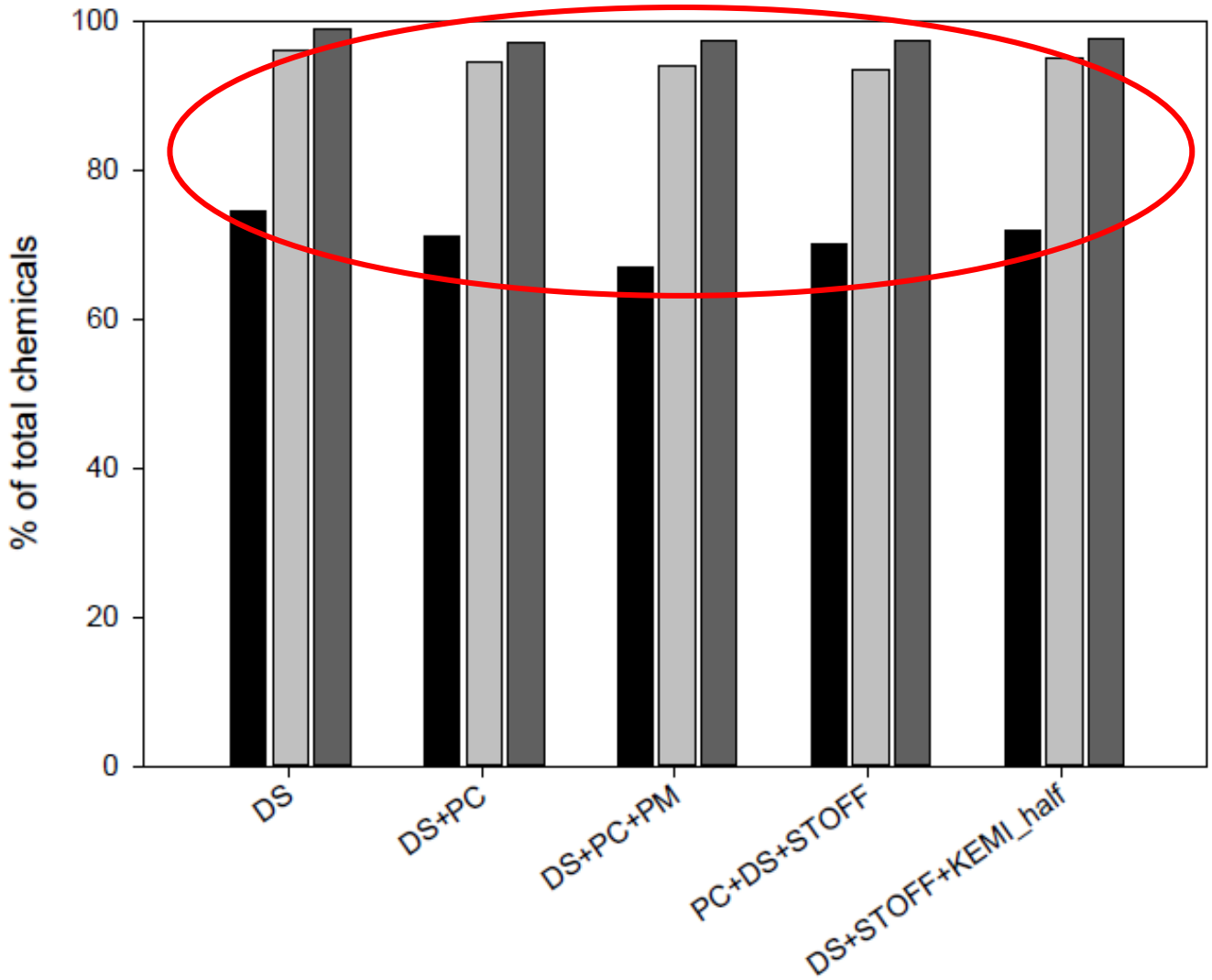


Candidate ranking using public resources

Using Metadata for Ranking

- Use available metadata to rank candidates
  - Associated data sources
    - Associated lists in the underlying database
    - Associated data sources in PubChem
    - Specific types (e.g. water, surfactants, pesticides etc.)
  - Number of associated literature articles (Pubmed)
  - **Chemicals in the environment** – the number of products/categories containing the chemical is a very important source of data

Identification ranks for 1783 chemicals using multiple data streams



DS: Data Sources
PC: PubChem
PM: PubMed
STOFF: DB
KEMI: DB

Is a bigger database better?

J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

RESEARCH ARTICLE

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

- 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?

SAME dataset for comparison

Compound class	Number in class	Average rank	Number of compounds in each position rank-ordered				
			#1	#2	#3	#4	#5+
Pharmaceutical drug	72	1.4	55	9	6	2	
Industrial chemicals	42	5.5	28	6	3		5
Personal care products	8	6.1	3	1			4
Steroid hormones	7	1.0	7				
Perfluorochemicals	6	1.2	5	1			
Pesticides	12	2.3	6	2	3		1
Veterinary drugs	3	1.3	2	1			
Dyes	2	1.0	2				
Food product/natural compounds	4	3.8	2			1	1
Illicit drugs	2	2.0	1		1		
Misc. molecules	3 [a]	1.3	2	1			

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

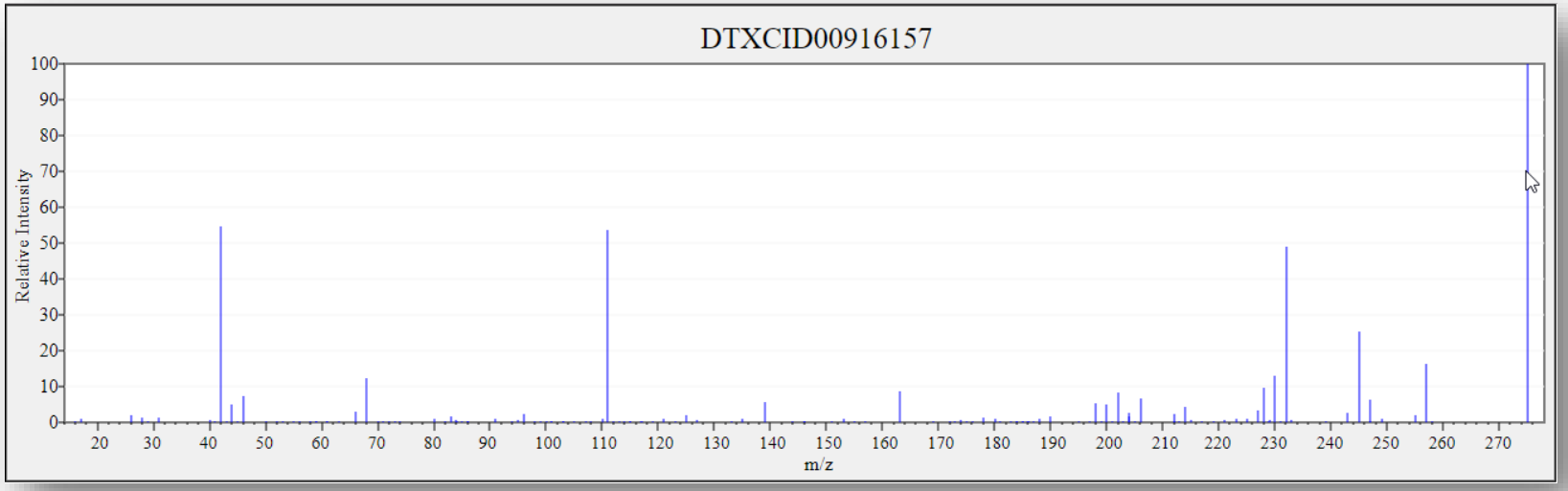
EPA United States Environmental Protection Agency

WORK IN PROGRESS

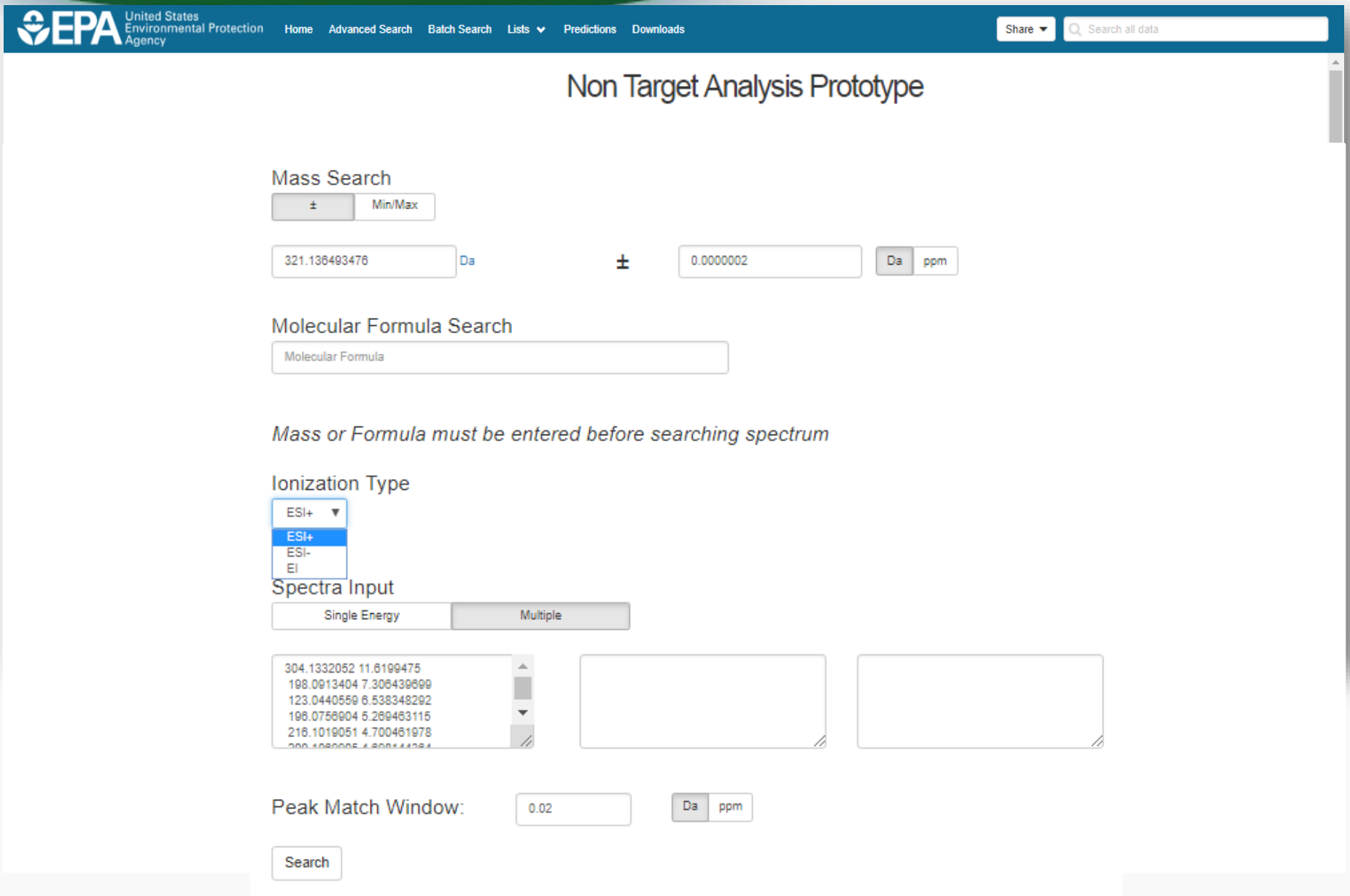
Predicted Mass Spectra

http://cfmid.wishartlab.com/

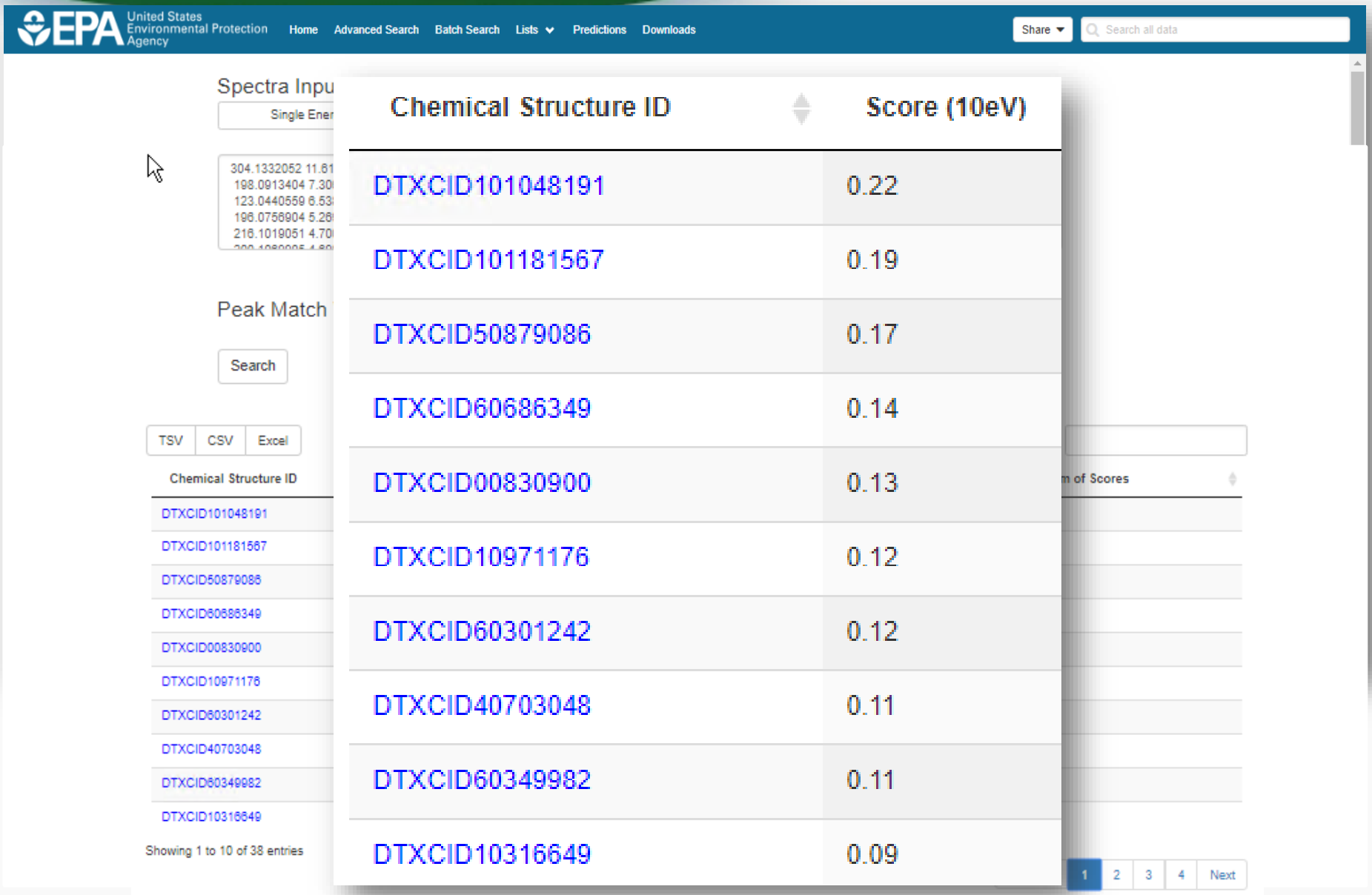
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra

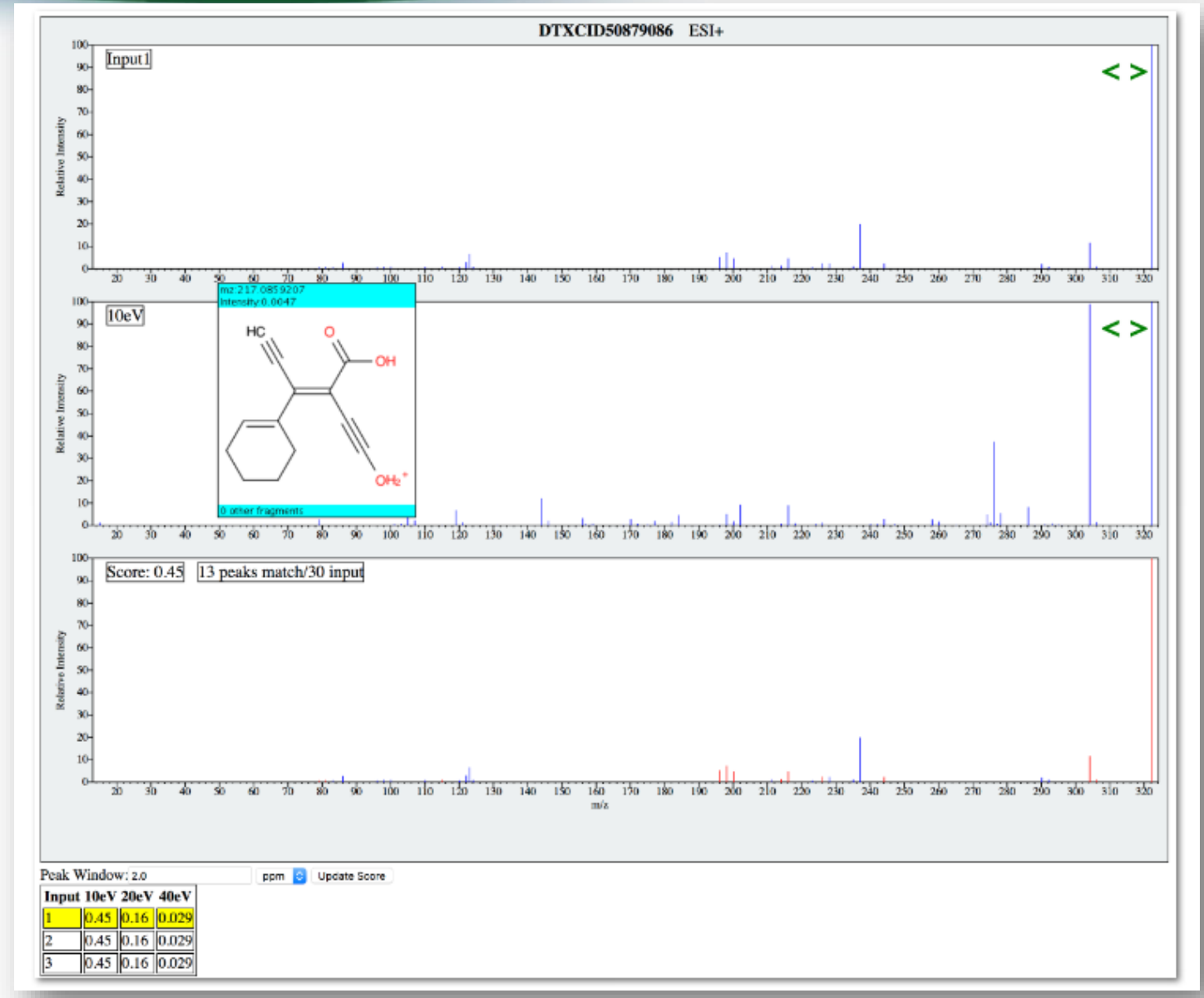


Search Expt. vs. Predicted Spectra

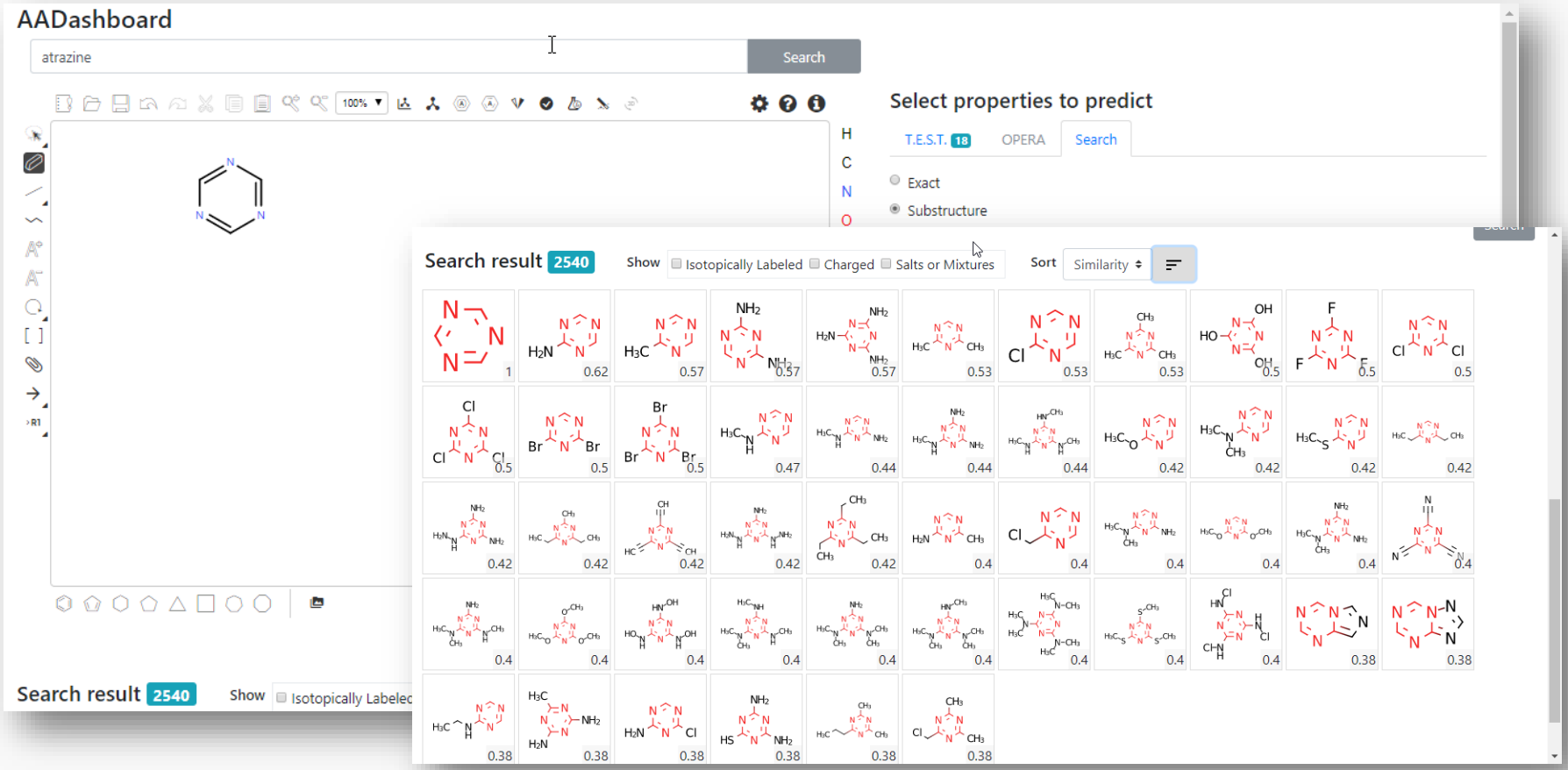


Chemical Structure ID	Score (10eV)
DTXCID101048191	0.22
DTXCID101181567	0.19
DTXCID50879086	0.17
DTXCID60686349	0.14
DTXCID00830900	0.13
DTXCID10971176	0.12
DTXCID60301242	0.12
DTXCID40703048	0.11
DTXCID60349982	0.11
DTXCID10316649	0.09

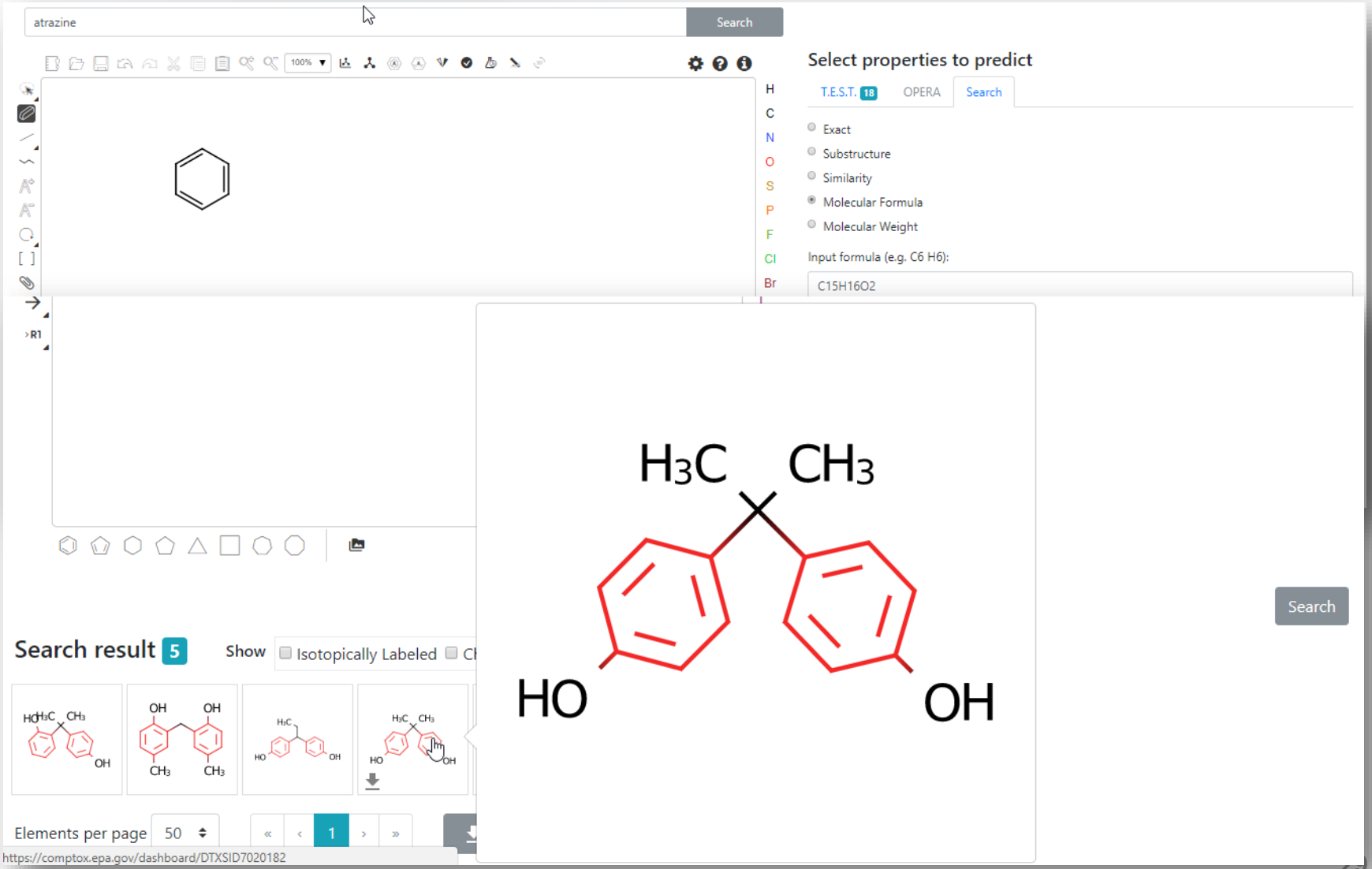
Spectral Viewer Comparison



Prototype Development



Prototype Development



Collision Cross Section Prediction

EPA United States Environmental Protection Agency

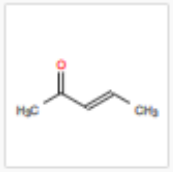
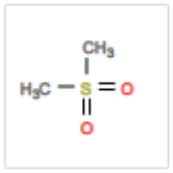
PNNL Collision Cross Section Database

Pacific Northwest NATIONAL LABORATORY

Search

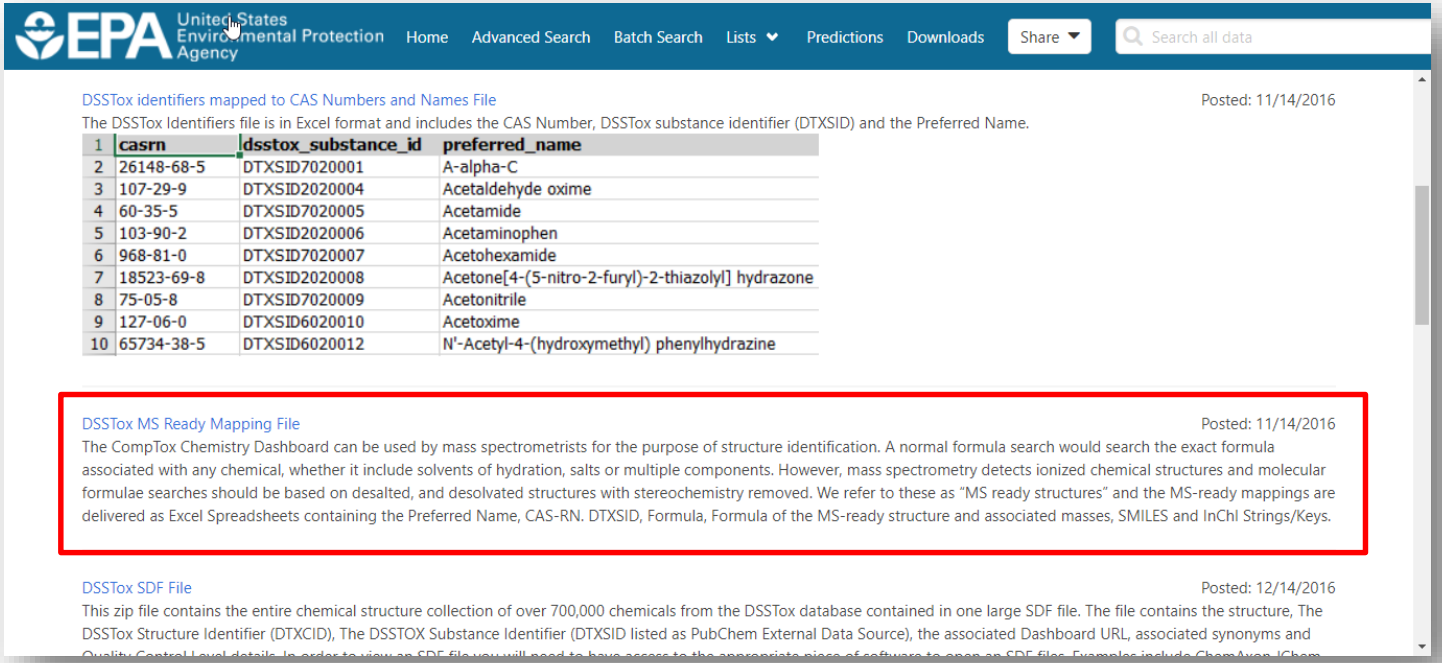
Showing 1 to 25 of 1000 rows 25 rows per page

‹ 1 2 3 4 5 ... 40 ›

Chemical	SMILES	InChI	Formula	Mass	CCS (Å^2)
 (3E)-pent-3-en-2-one	*	*	C_5H_8O	84.0575	[M-H]$^-$ 112.1 ISiCLE Lite v0.1.0 [M+Na]$^+$ 112.6 ISiCLE Lite v0.1.0 [M+H]$^+$ 113.1 ISiCLE Lite v0.1.0
 Dimethyl sulfone	*	*	$C_2H_6O_2S$	94.0089	[M-H]$^-$ 106.9 ISiCLE Lite v0.1.0 [M+Na]$^+$ 107.3 ISiCLE Lite v0.1.0 [M+H]$^+$ 108.1 ISiCLE Lite v0.1.0
 isothiocyanatocyclopropane	*	*	C_4H_5NS	99.0143	[M-H]$^-$ 111.9 ISiCLE Lite v0.1.0 [M+Na]$^+$ 112.1 ISiCLE Lite v0.1.0 [M+H]$^+$ 110.0 ISiCLE Lite v0.1.0

API services and Open Data

- Groups waiting on our API and web services
- Mass Spec companies instrument integration
- Release will be in iterations but for now our data are available



EPA United States Environmental Protection Agency

Benefiting the community with Open Data

NORMAN Suspect List Exchange

https://www.norman-network.com/?q=node/236

EPA United States Environmental Protection Agency

NORMAN

Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances

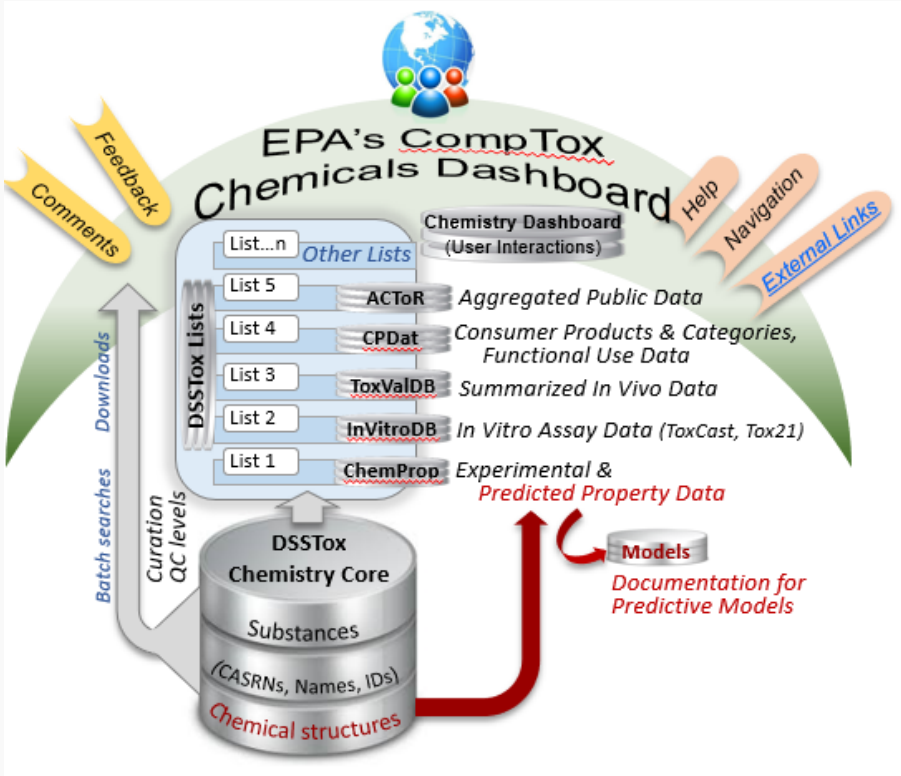
Wastewater Suspect List based on Swedish Product Data	Wastewater Suspect List **Original File with Mapped DTXSIDs** (12/02/2019)	KEMIWWSUS **InChIKeys** (12/02/2019)	A prioritized list of 1,123 substances relevant for wastewater based on Swedish product registry data, including scores. Provided by Stellan Fischer, KEMI.
Algal toxins list from CompTox	ALGALTOX **XLSX**, **CSV** (14/02/2019) CompTox **ALGALTOX List**	ALGALTOX **InChIKeys** (14/02/2019)	List of algal toxins (generated during blooms) from the CompTox Chemicals Dashboard.
CCL 4 Chemical Candidate List	CCL4 **XLSX**, **CSV** (14/02/2019) CompTox **CCL4 List**	CCL4 **InChIKeys** (14/02/2019)	Contaminants that are not (yet) regulated in the USA but are known or anticipated to occur in public water systems; from CompTox.
Hydrogen Deuterium Exchange (HDX) Standard Set	HDXNOEX **XLSX**, **CSV** (14/02/2019) CompTox **HDXNOEX List** CompTox **HDXEXCH List**	HDXNOEX **InChIKeys** (14/02/2019)	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule HRMS (Ruttkies et al. submitted). **HDXEXCH** list also contains observed deuterated species.
Neurotoxicants Collection from Public Resources	NEUROTOXINS **XLSX**, **CSV** (14/02/2019) CompTox **NEUROTOXINS List**	NEUROTOXINS **InChIKeys** (14/02/2019)	A list of neurotoxicants compiled from public resources, details on CompTox and Schymanski *et al.* (submitted).
Statins Collection from Public Resources	STATINS **XLSX**, **CSV** (14/02/2019) CompTox **STATINS List**	STATINS **InChIKeys** (14/02/2019)	A list of statins (lipid-lowering medications) compiled from public resources, details on CompTox.
Synthetic Cannabinoids and Psychoactive Compounds	SYNTHCANNAB **XLSX**, **CSV** (14/02/2019) CompTox **SYNTHCANNAB List**	SYNTHCANNAB **InChIKeys** (14/02/2019)	A list of synthetic cannabinoids and psychoactive compounds assembled from public resources, from CompTox.

Integration to MetFrag in place

https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0299-2

Conclusion

- Dashboard access to data for ~875,000 chemicals
- “MS-Ready” structures dramatically enable searches
- Ongoing list curation expands data access
- Candidate ranking supported by integrated meta-data
- Integration of predicted spectra (ESI+/- and EI) will assist candidate ranking
- API and services will be delivered



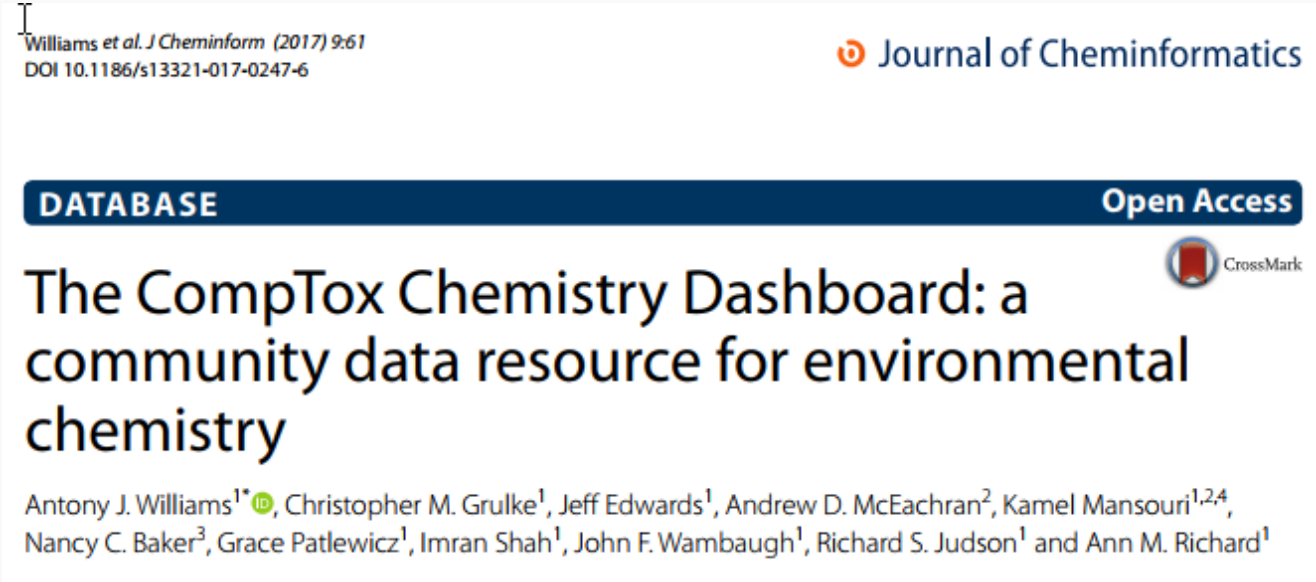
Acknowledgements

EPA-RTP

- *An enormous team of contributors from NCCT, especially the IT software development team*
- *Our curation team for their care and focus on data quality*
- *Multiple centers and laboratories across the EPA*
- *Many public domain databases and open data contributors*

Contact

Antony Williams
NCCT, US EPA Office of Research and Development,
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821



Williams et al. J Cheminform (2017) 9:61
DOI 10.1186/s13321-017-0247-6

Journal of Cheminformatics

DATABASE Open Access

The CompTox Chemistry Dashboard: a community data resource for environmental chemistry

Antony J. Williams[1*], Christopher M. Grulke[1], Jeff Edwards[1], Andrew D. McEachran[2], Kamel Mansouri[1,2,4], Nancy C. Baker[3], Grace Patlewicz[1], Imran Shah[1], John F. Wambaugh[1], Richard S. Judson[1] and Ann M. Richard[1]

https://doi.org/10.1186/s13321-017-0247-6