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

Non-targeted analysis supported by data and cheminformatics delivered via the CompTox Chemicals Dashboard

***Antony Williams**[1], Alex Chao[2], Tom Transue[3], Tommy Cathey[3], Elin Ulrich[1] and Jon Sobus[1]*

1) Center for Computational Toxicology and Exposure, U.S. Environmental Protection Agency, RTP, NC
2) Oak Ridge Institute of Science and Education (ORISE) Research Participant, RTP, NC
3) GDIT, Research Triangle Park, North Carolina, United State

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

November 2019
SETAC, Toronto, Canada

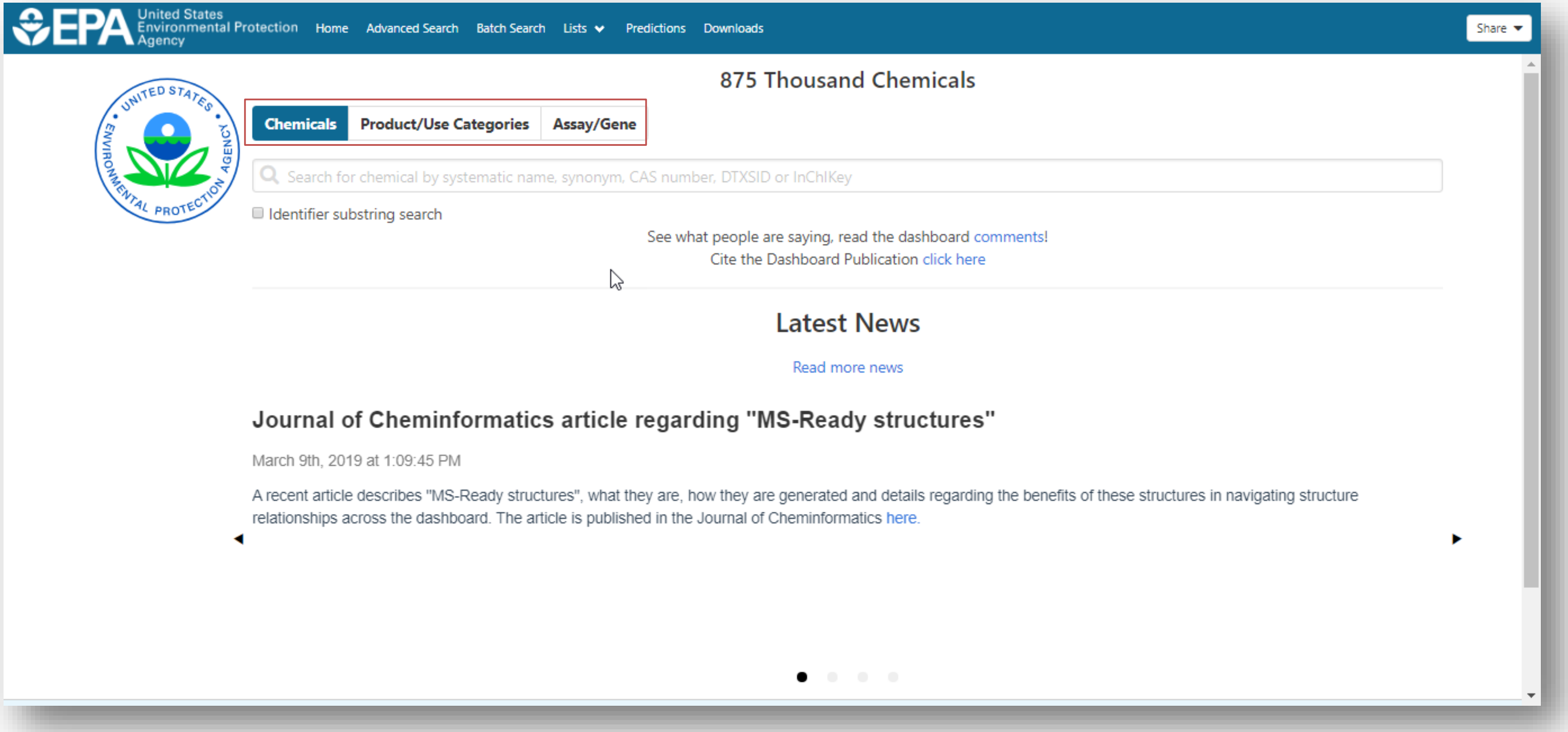
An intro to the Dashboard

- Freely available web-based database from the National Center for Computational Toxicology
- Providing data for 875,000 substances including
  - Experimental and predicted physicochemical properties
  - *In vivo* toxicity data harvested from dozens of public resources
  - *In vitro* bioactivity data for thousands of chemicals and assays
  - Exposure data including chemicals in consumer products
  - Real time predictions for >20 physchem and toxicological endpoints
- Dashboard is used by mass spectrometrists for chemical identification
- A quick view of general capabilities…

CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard

875k Chemical Substances



EPA United States Environmental Protection Agency
Home
Advanced Search
Batch Search
Lists
Predictions
Downloads
Share
Chemicals
Product/Use Categories
Assay/Gene
Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey
□ Identifier substring search
Read more news

875 Thousand Chemicals

See what people are saying, read the dashboard comments!

Cite the Dashboard Publication click here

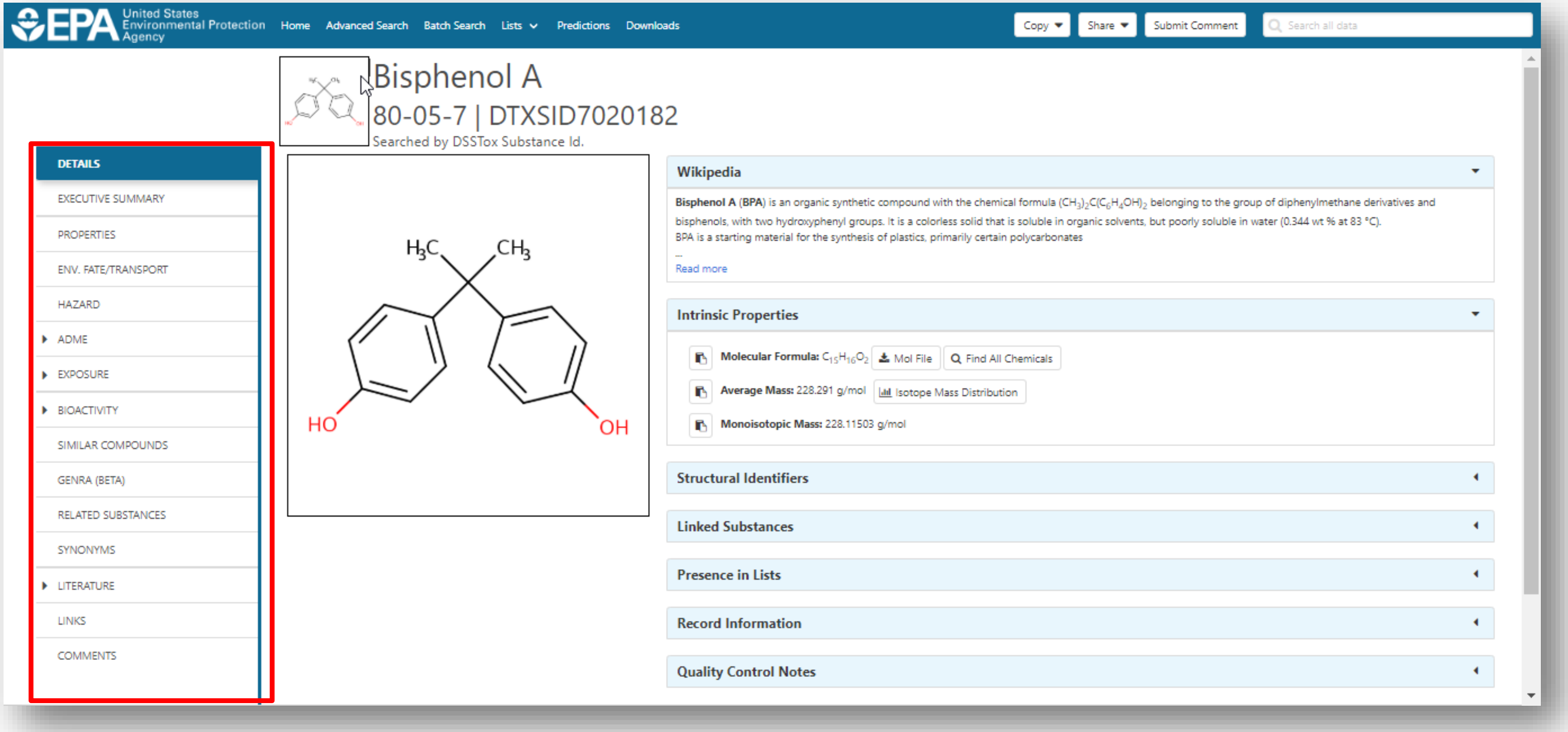
Latest News

Journal of Cheminformatics article regarding "MS-Ready structures"

March 9th, 2019 at 1:09:45 PM

A recent article describes "MS-Ready structures", what they are, how they are generated and details regarding the benefits of these structures in navigating structure relationships across the dashboard. The article is published in the Journal of Cheminformatics here.

Detailed Chemical Pages



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Wikipedia

Bisphenol A (**BPA**) is an organic synthetic compound with the chemical formula $(CH_3)_2C(C_6H_4OH)_2$ belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water (0.344 wt % at 83 °C).
BPA is a starting material for the synthesis of plastics, primarily certain polycarbonates

...

Read more

Intrinsic Properties

Molecular Formula: $C_{15}H_{16}O_2$

Average Mass: 228.291 g/mol

Monoisotopic Mass: 228.11503 g/mol

Structural Identifiers

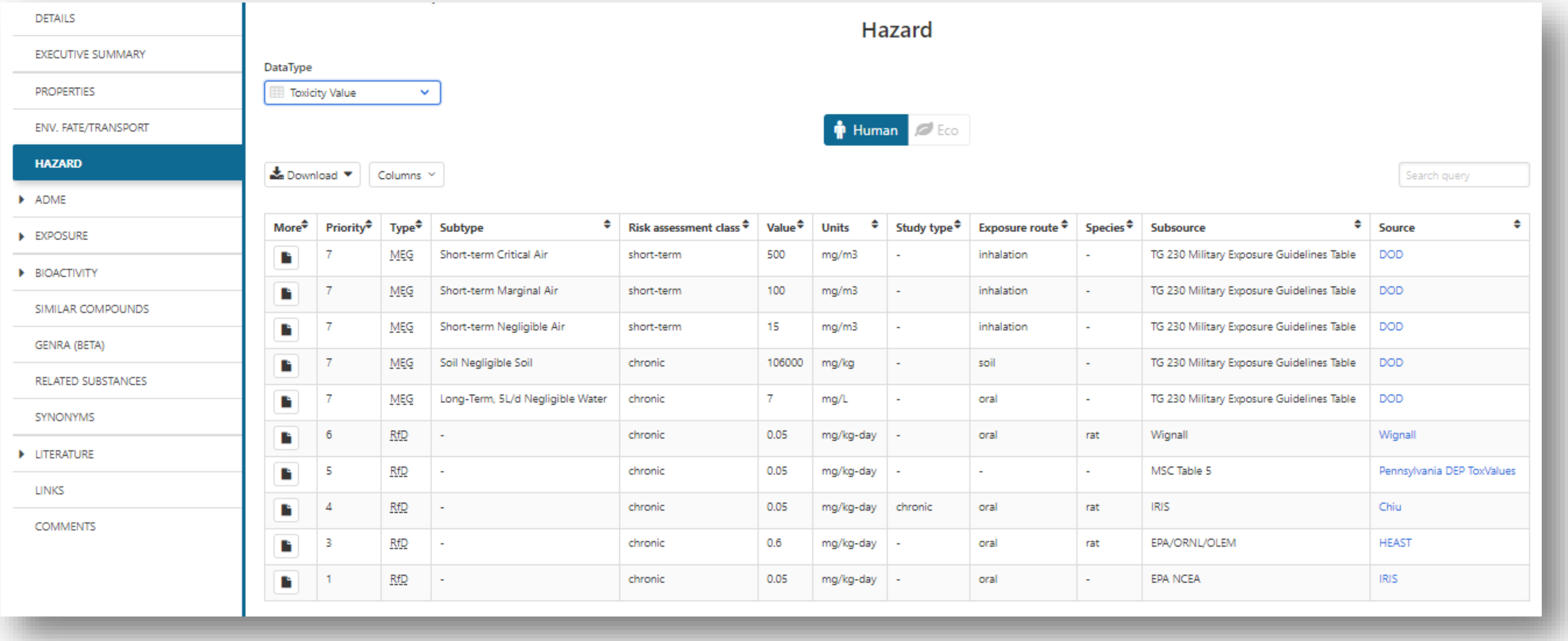
Linked Substances

Presence in Lists

Record Information

Quality Control Notes

Access to Chemical Hazard Data

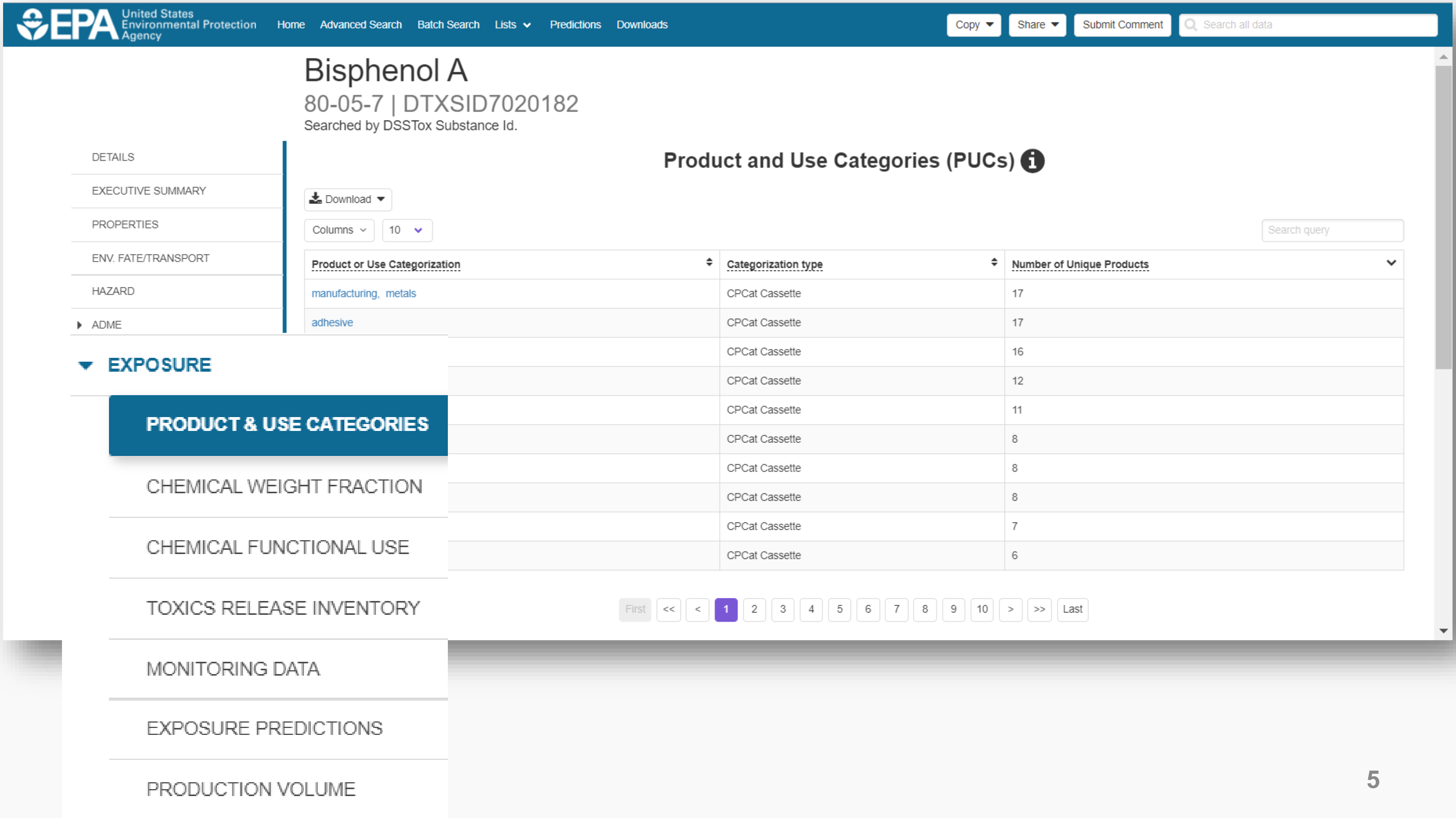


Hazard

DataType: Toxicity Value

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

Sources of Exposure to Chemicals



Bisphenol A

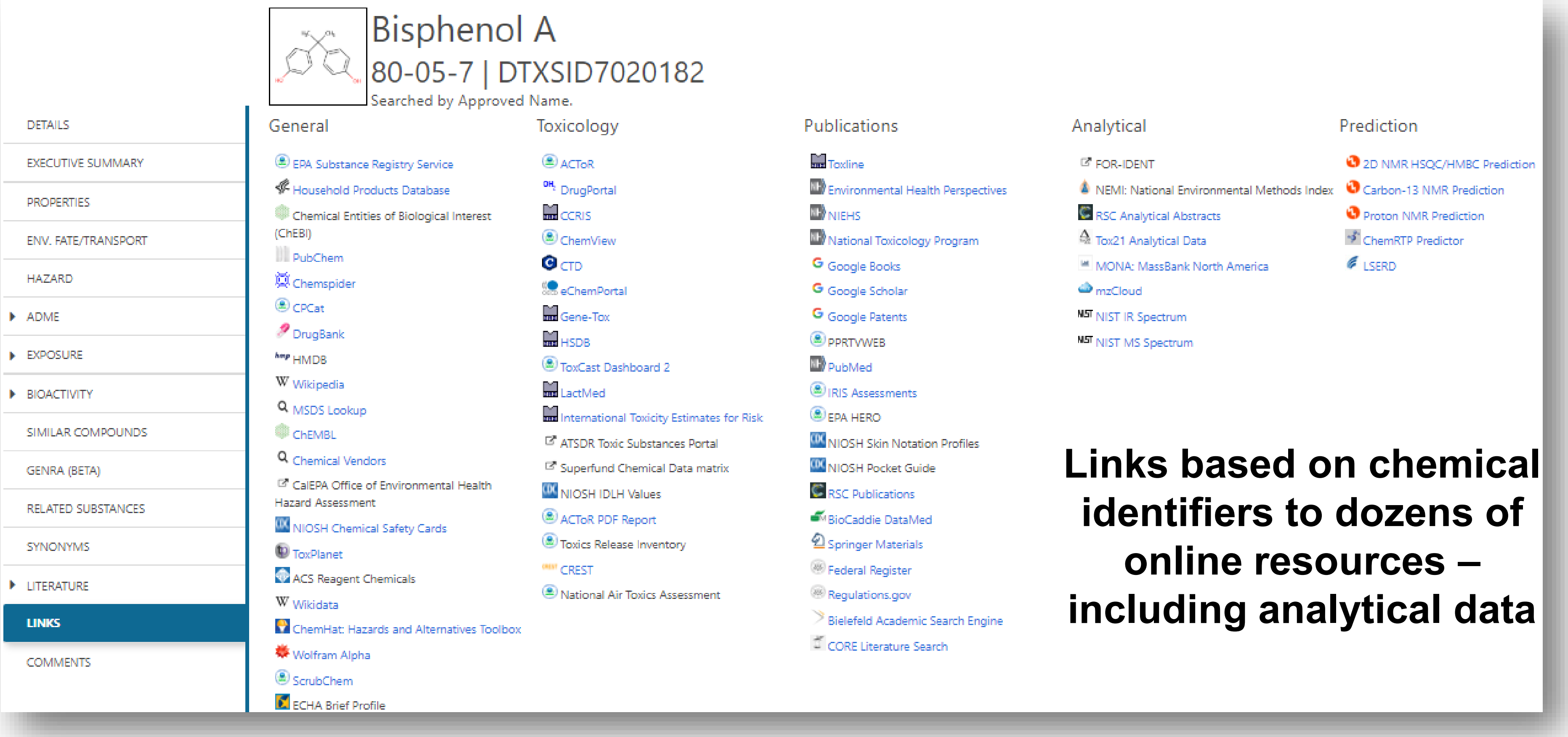
80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Product and Use Categories (PUCs)

Product or Use Categorization	Categorization type	Number of Unique Products
manufacturing, metals	CPCat Cassette	17
adhesive	CPCat Cassette	17
	CPCat Cassette	16
	CPCat Cassette	12
	CPCat Cassette	11
	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	7
	CPCat Cassette	6

Link Access



Bisphenol A

80-05-7 | DTXSID7020182

Searched by Approved Name.

General

- EPA Substance Registry Service
- Household Products Database
- Chemical Entities of Biological Interest (ChEBI)
- PubChem
- Chemspider
- CPCat
- DrugBank
- HMDB
- Wikipedia
- MSDS Lookup
- ChEMBL
- Chemical Vendors
- CalEPA Office of Environmental Health Hazard Assessment
- NIOSH Chemical Safety Cards
- ToxPlanet
- ACS Reagent Chemicals
- Wikidata
- ChemHat: Hazards and Alternatives Toolbox
- Wolfram Alpha
- ScrubChem
- ECHA Brief Profile

Toxicology

- ACToR
- DrugPortal
- CCRIS
- ChemView
- CTD
- eChemPortal
- Gene-Tox
- HSDB
- ToxCast Dashboard 2
- LactMed
- International Toxicity Estimates for Risk
- ATSDR Toxic Substances Portal
- Superfund Chemical Data matrix
- NIOSH IDLH Values
- ACToR PDF Report
- Toxics Release Inventory
- CREST
- National Air Toxics Assessment

Publications

- Toxline
- Environmental Health Perspectives
- NIEHS
- National Toxicology Program
- Google Books
- Google Scholar
- Google Patents
- PPRTVWEB
- PubMed
- IRIS Assessments
- EPA HERO
- NIOSH Skin Notation Profiles
- NIOSH Pocket Guide
- RSC Publications
- BioCaddie DataMed
- Springer Materials
- Federal Register
- Regulations.gov
- Bielefeld Academic Search Engine
- CORE Literature Search

Analytical

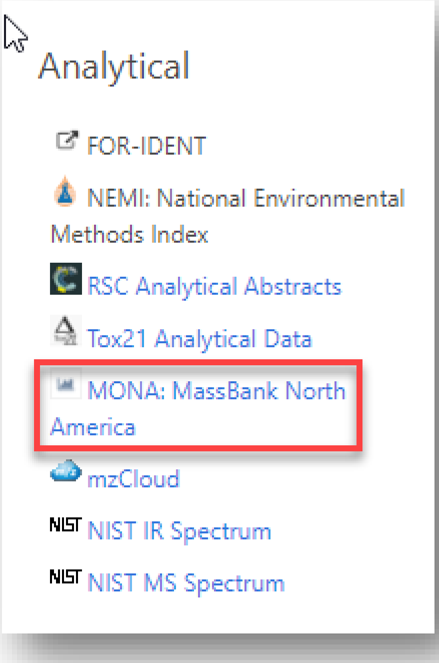
- FOR-IDENT
- NEMI: National Environmental Methods Index
- RSC Analytical Abstracts
- Tox21 Analytical Data
- MONA: MassBank North America
- mzCloud
- NIST IR Spectrum
- NIST MS Spectrum

Prediction

- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction
- ChemRTP Predictor
- LSERD

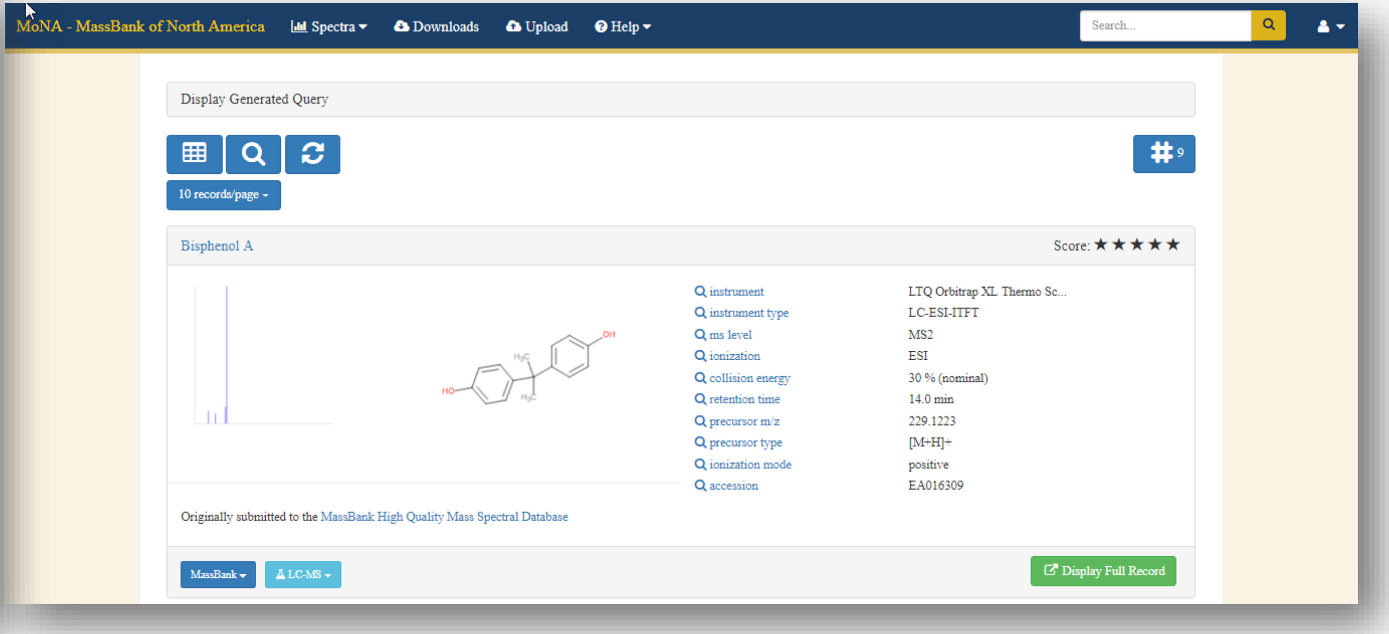
Links based on chemical identifiers to dozens of online resources – including analytical data

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



Analytical

- FOR-IDENT
- NEMI: National Environmental Methods Index
- RSC Analytical Abstracts
- Tox21 Analytical Data
- MONA: MassBank North America
- mzCloud
- NIST IR Spectrum
- NIST MS Spectrum



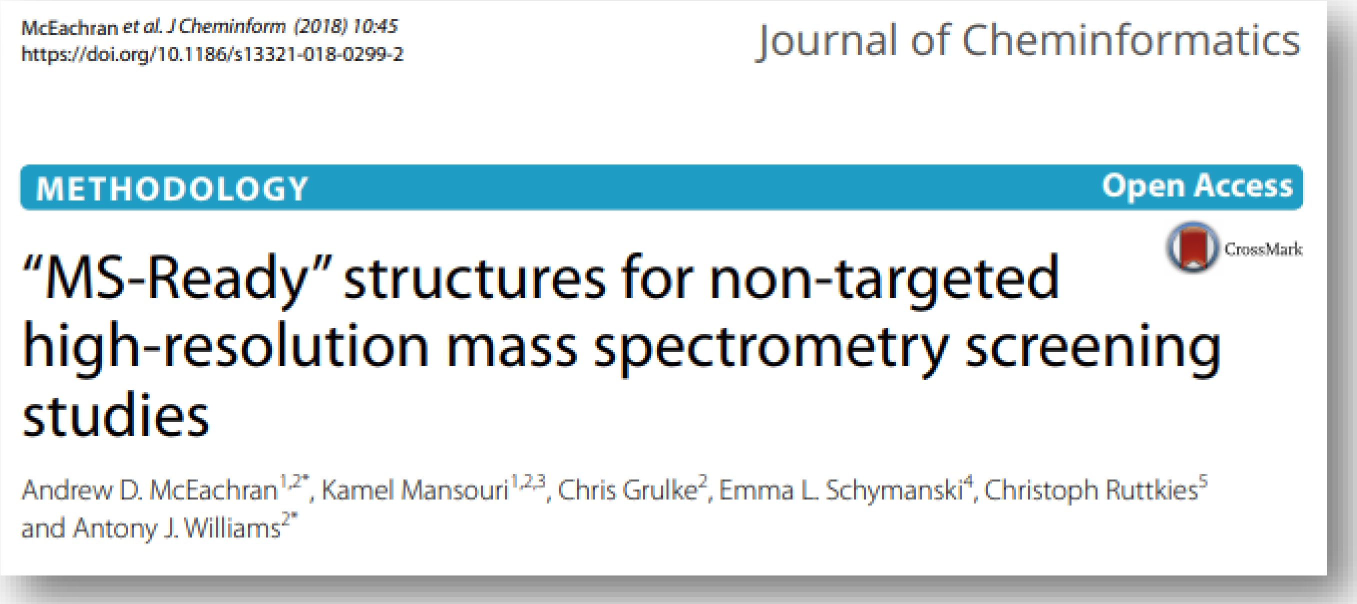
Bisphenol A

Score: ★★★★★

instrument	LTQ Orbitrap XL Thermo Sc...
instrument type	LC-ESI-ITFT
ms level	MS2
ionization	ESI
collision energy	30 % (nominal)
retention time	14.0 min
precursor m/z	229.1223
precursor type	[M+H]+
ionization mode	positive
accession	EA016309

Originally submitted to the MassBank High Quality Mass Spectral Database

"MS-ready" structures



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

Journal of Cheminformatics

METHODOLOGY Open Access

CrossMark

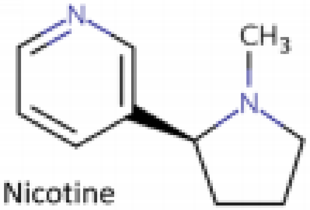
"MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies

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*]

Overview of MS-Ready Structures

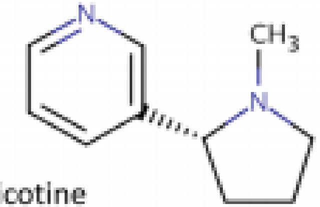
- All structure-based chemical substances are algorithmically processed to
  - Split multicomponent chemicals into individual structures
  - Desalt and neutralize individual structures
  - Remove stereochemical bonds from all chemicals

LEGEND: Name, SMILES
DTXSID | InChIKey 1st Block
CAS | **Monoiso. Mass** | logP | **Sources**
Data on: **Toxicity** | **Exposure** | **Bioassays**



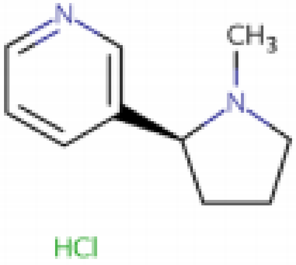
Nicotine

CN1CCC[C@H]1C1=CN=CC=C1
DTXSID1020930| SNICXCGAKADSCV
54-11-5 | **162.1157**| 0.929| **72**
Tox: **yes**| Expo: **yes**| Bioassay: **yes**



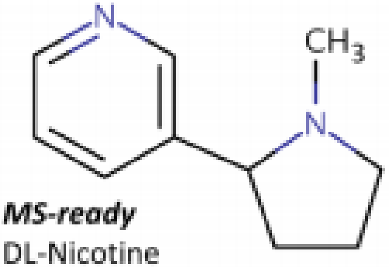
D-Nicotine

CN1CCC[C@@H]1C1=CN=CC=C1
DTXSID004635| SNICXCGAKADSCV
25162-00-9 | **162.1157**| 0.929| **20**
Tox: **no**| Expo: **yes**| Bioassay: **yes**



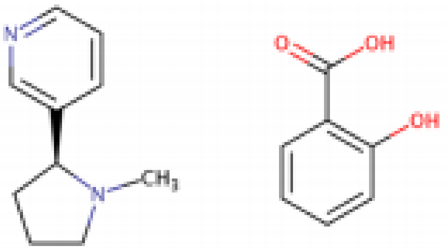
Nicotine hydrochloride

Cl.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID602093| HDJBTCAJIMNXEW
2820-51-1 | **198.0924** | 0.929| **9**
Tox: **no**| Expo: **yes**| Bioassay: **yes**



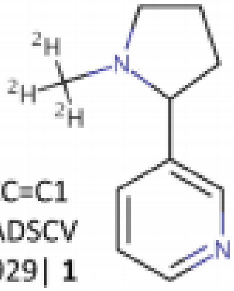
MS-ready
DL-Nicotine

CN1CCCC1C1=CN=CC=C1
DTXSID3048154 | SNICXCGAKADSCV
22083-74-5 | **162.1157**| 0.953| 9
Tox: **yes**| Expo: **no**| Bioassay: **yes**



Benzoic acid, 2-hydroxy-, compd. with
3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)

OC(=O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID5075319| AIBWPBUAKCMKNS
29790-52-1| **300.1474**| 0.929| **6**
Tox: **no**| Expo: **yes**| Bioassay: **no**



DL-Nicotine-d3

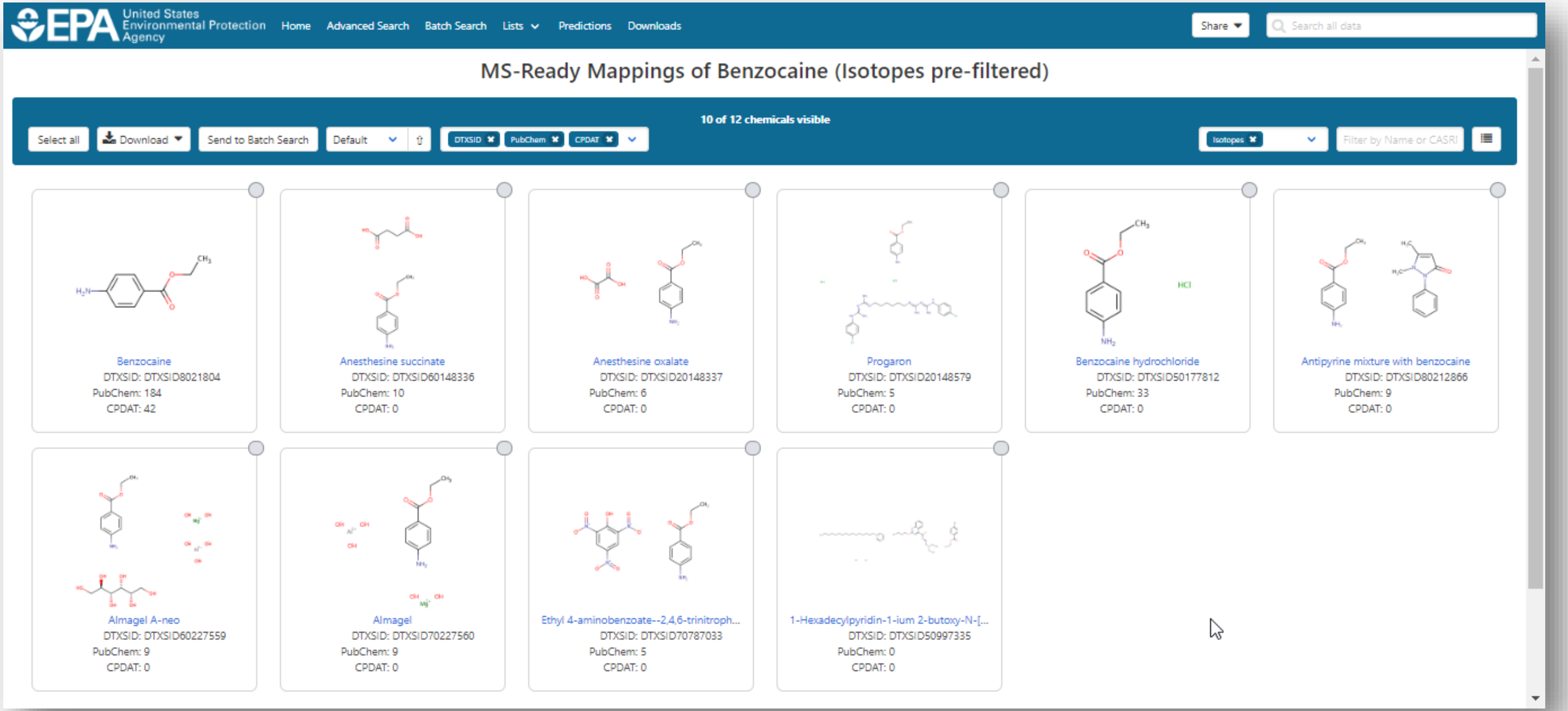
[2H]C([2H])([2H])N1CCCC1C1=CN=CC=C1
DTXSID80442666| SNICXCGAKADSCV
69980-24-1| **165.1345**| 0.929| **1**
Tox: **no**| Expo: **no**| Bioassay: **no**

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 Set
All substances containing component



MS-Ready Mappings of Benzocaine (Isotopes pre-filtered)

10 of 12 chemicals visible

Benzocaine
DTXSID: DTXSID8021804
PubChem: 184
CPDAT: 42

Anesthesine succinate
DTXSID: DTXSID60148336
PubChem: 10
CPDAT: 0

Anesthesine oxalate
DTXSID: DTXSID20148337
PubChem: 6
CPDAT: 0

Progaron
DTXSID: DTXSID20148579
PubChem: 5
CPDAT: 0

Benzocaine hydrochloride
DTXSID: DTXSID50177812
PubChem: 33
CPDAT: 0

Antipyrine mixture with benzocaine
DTXSID: DTXSID80212866
PubChem: 9
CPDAT: 0

Almagel A-neo
DTXSID: DTXSID60227559
PubChem: 9
CPDAT: 0

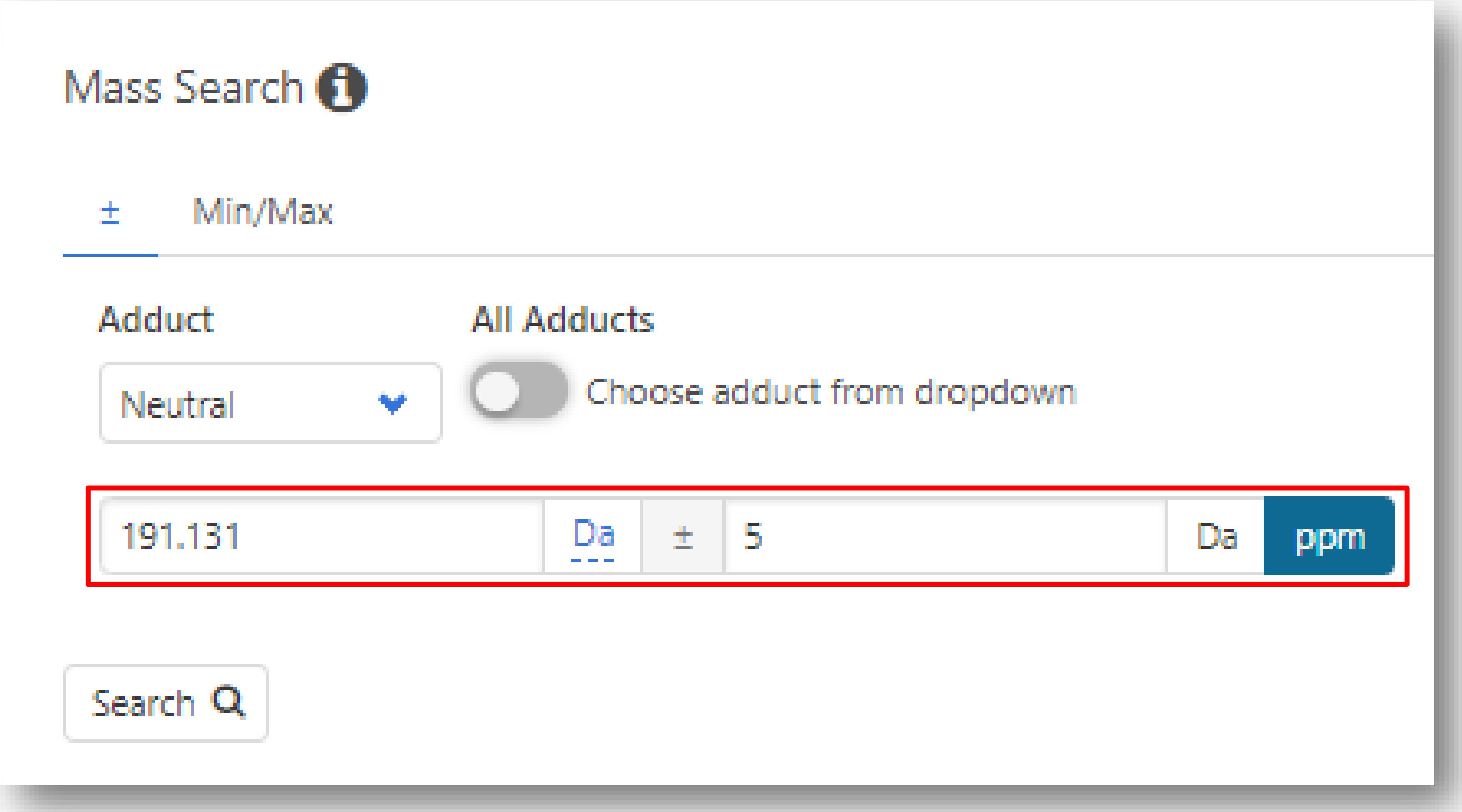
Almagel
DTXSID: DTXSID70227560
PubChem: 9
CPDAT: 0

Ethyl 4-aminobenzoate--2,4,6-trinitroph...
DTXSID: DTXSID70787033
PubChem: 5
CPDAT: 0

1-Hexadecylpyridin-1-ium 2-butoxy-N-[...
DTXSID: DTXSID50997335
PubChem: 0
CPDAT: 0

Mass/Formula Searching and Metadata Ranking

Advanced Searches
Mass Search



Mass Search

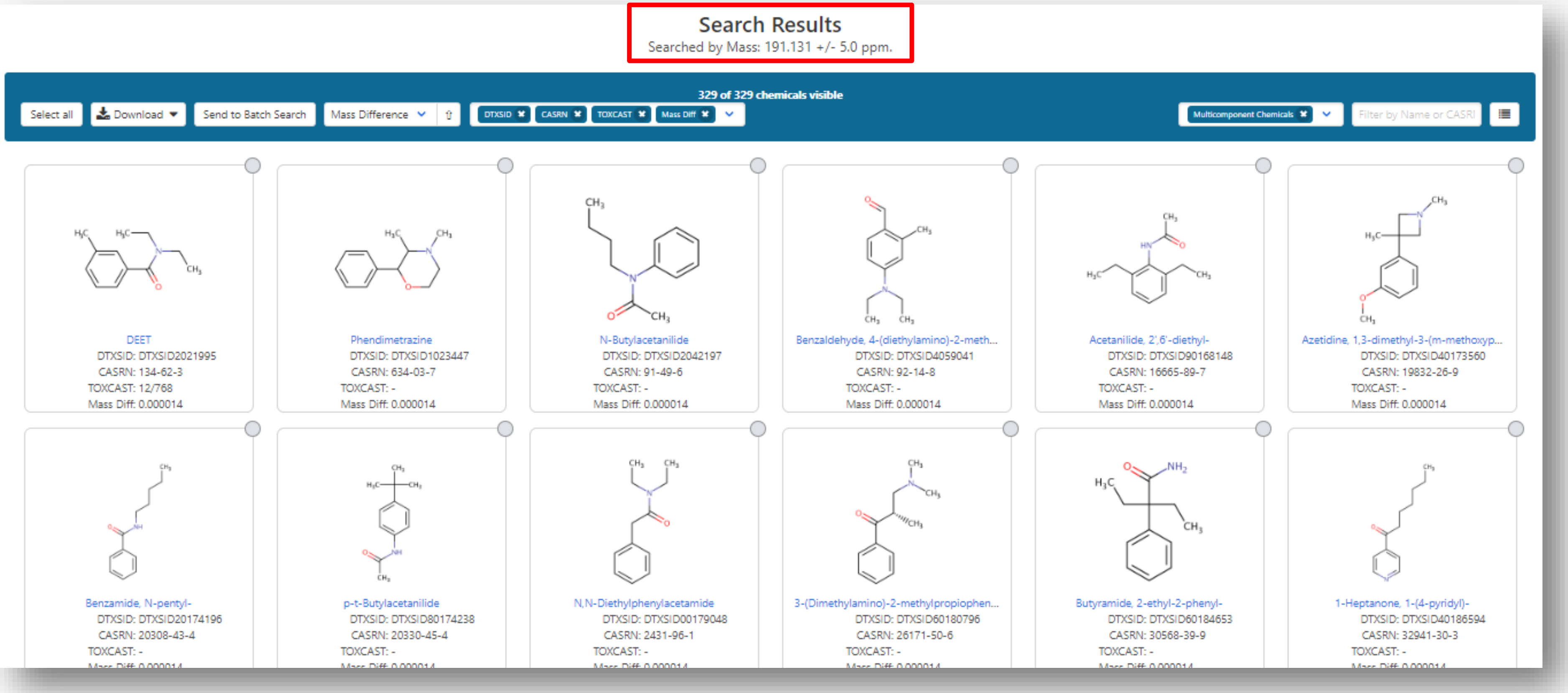
± Min/Max

Adduct: Neutral

All Adducts: ○ Choose adduct from dropdown

191.131 Da ± 5 Da ppm

Advanced Searches
Mass Search



Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

329 of 329 chemicals visible

DEET
DTXSID: DTXSID2021995
CASRN: 134-62-3
TOXCAST: 12/768
Mass Diff: 0.000014

Phendimetrazine
DTXSID: DTXSID1023447
CASRN: 634-03-7
TOXCAST: -
Mass Diff: 0.000014

N-Butylacetanilide
DTXSID: DTXSID2042197
CASRN: 91-49-6
TOXCAST: -
Mass Diff: 0.000014

Benzaldehyde, 4-(diethylamino)-2-meth...
DTXSID: DTXSID4059041
CASRN: 92-14-8
TOXCAST: -
Mass Diff: 0.000014

Acetanilide, 2',6'-diethyl-
DTXSID: DTXSID90168148
CASRN: 16665-89-7
TOXCAST: -
Mass Diff: 0.000014

Azetidine, 1,3-dimethyl-3-(m-methoxyp...
DTXSID: DTXSID40173560
CASRN: 19832-26-9
TOXCAST: -
Mass Diff: 0.000014

Benzamide, N-pentyl-
DTXSID: DTXSID20174196
CASRN: 20308-43-4
TOXCAST: -
Mass Diff: 0.000014

p-t-Butylacetanilide
DTXSID: DTXSID80174238
CASRN: 20330-45-4
TOXCAST: -
Mass Diff: 0.000014

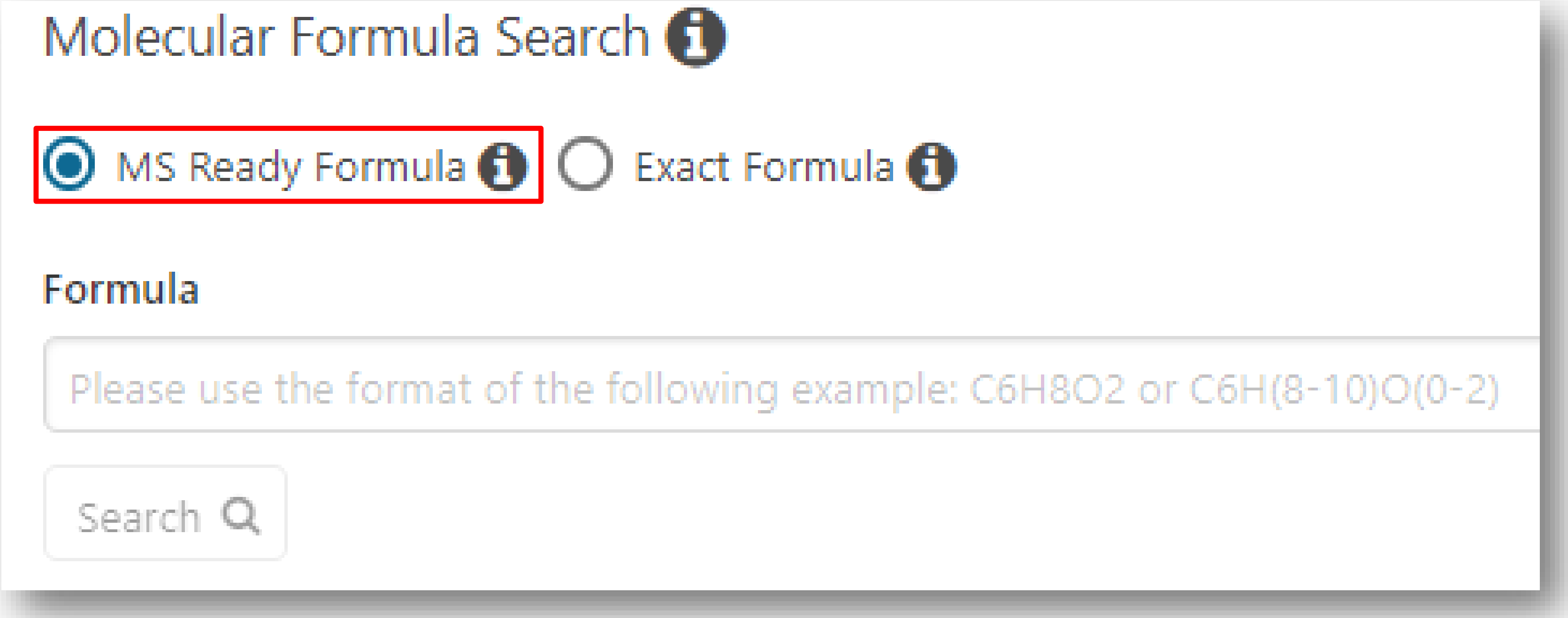
N,N-Diethylphenylacetamide
DTXSID: DTXSID00179048
CASRN: 2431-96-1
TOXCAST: -
Mass Diff: 0.000014

3-(Dimethylamino)-2-methylpropiophen...
DTXSID: DTXSID60180796
CASRN: 26171-50-6
TOXCAST: -
Mass Diff: 0.000014

Butyramide, 2-ethyl-2-phenyl-
DTXSID: DTXSID60184653
CASRN: 30568-39-9
TOXCAST: -
Mass Diff: 0.000014

1-Heptanone, 1-(4-pyridyl)-
DTXSID: DTXSID40186594
CASRN: 32941-30-3
TOXCAST: -
Mass Diff: 0.000014

MS-Ready Structures for **Formula Search**



Molecular Formula Search

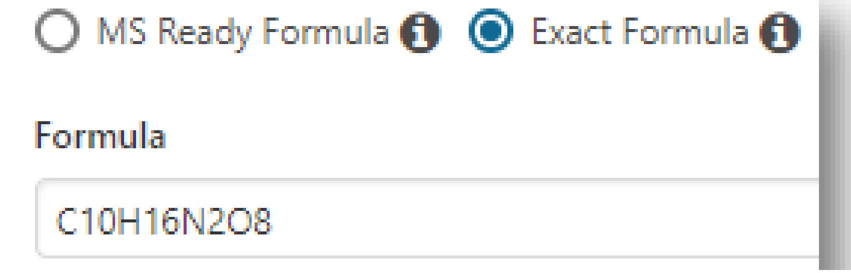
● MS Ready Formula ○ Exact Formula

Formula

Please use the format of the following example: C6H8O2 or C6H(8-10)O(0-2)

MS-Ready Mappings

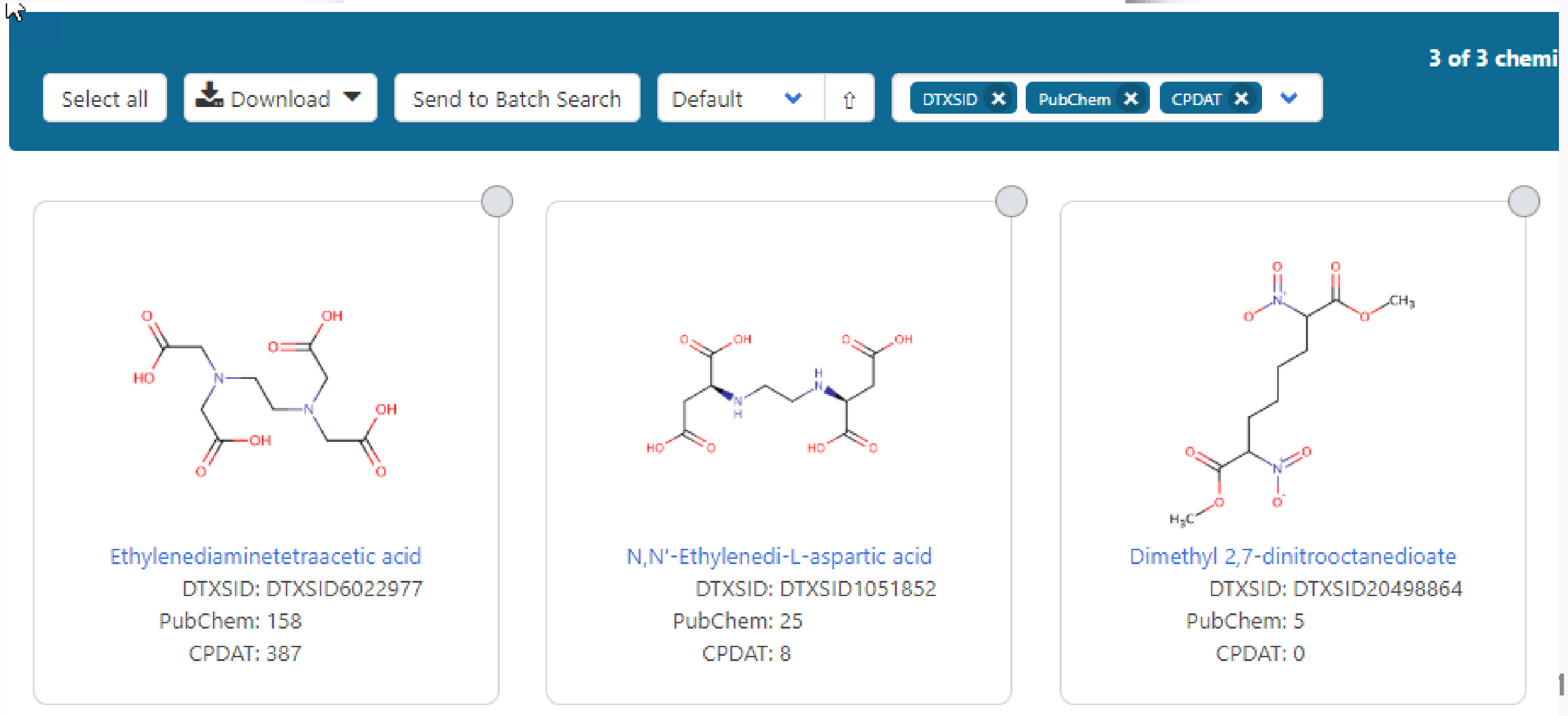
- **EXACT Formula**: C10H16N2O8: **3** Hits



○ MS Ready Formula ● Exact Formula

Formula

C10H16N2O8



3 of 3 chemi

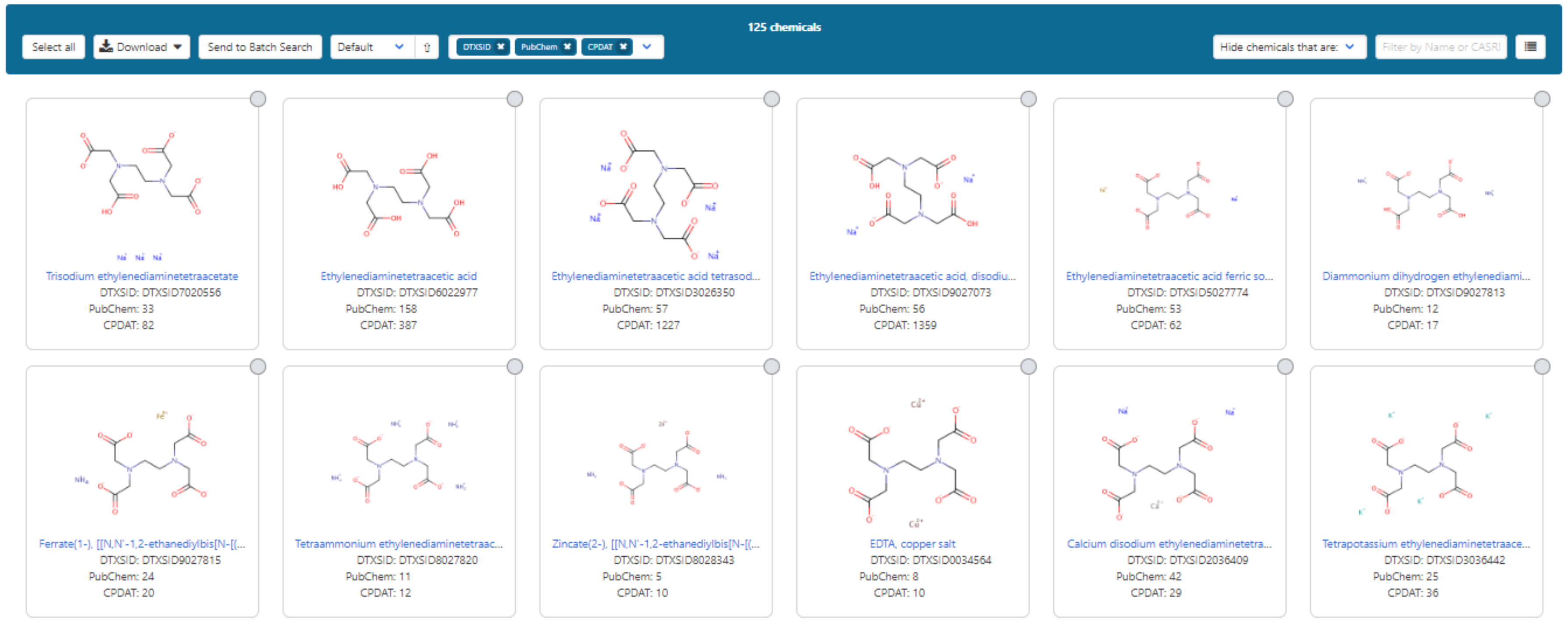
Ethylenediaminetetraacetic acid
DTXSID: DTXSID6022977
PubChem: 158
CPDAT: 387

N,N'-Ethylenedi-L-aspartic acid
DTXSID: DTXSID1051852
PubChem: 25
CPDAT: 8

Dimethyl 2,7-dinitrooctanedioate
DTXSID: DTXSID20498864
PubChem: 5
CPDAT: 0

MS-Ready Mappings

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



125 chemicals

Trisodium ethylenediaminetetraacetate
DTXSID: DTXSID7020556
PubChem: 33
CPDAT: 82

Ethylenediaminetetraacetic acid
DTXSID: DTXSID6022977
PubChem: 158
CPDAT: 387

Ethylenediaminetetraacetic acid tetrasod...
DTXSID: DTXSID3026350
PubChem: 57
CPDAT: 1227

Ethylenediaminetetraacetic acid, disodiu...
DTXSID: DTXSID9027073
PubChem: 56
CPDAT: 1359

Ethylenediaminetetraacetic acid ferric so...
DTXSID: DTXSID5027774
PubChem: 53
CPDAT: 62

Diammonium dihydrogen ethylenediami...
DTXSID: DTXSID9027813
PubChem: 12
CPDAT: 17

Ferrate(1-), [[N,N'-1,2-ethanediylbis[N-[(...
DTXSID: DTXSID9027815
PubChem: 24
CPDAT: 20

Tetraammonium ethylenediaminetetraac...
DTXSID: DTXSID8027820
PubChem: 11
CPDAT: 12

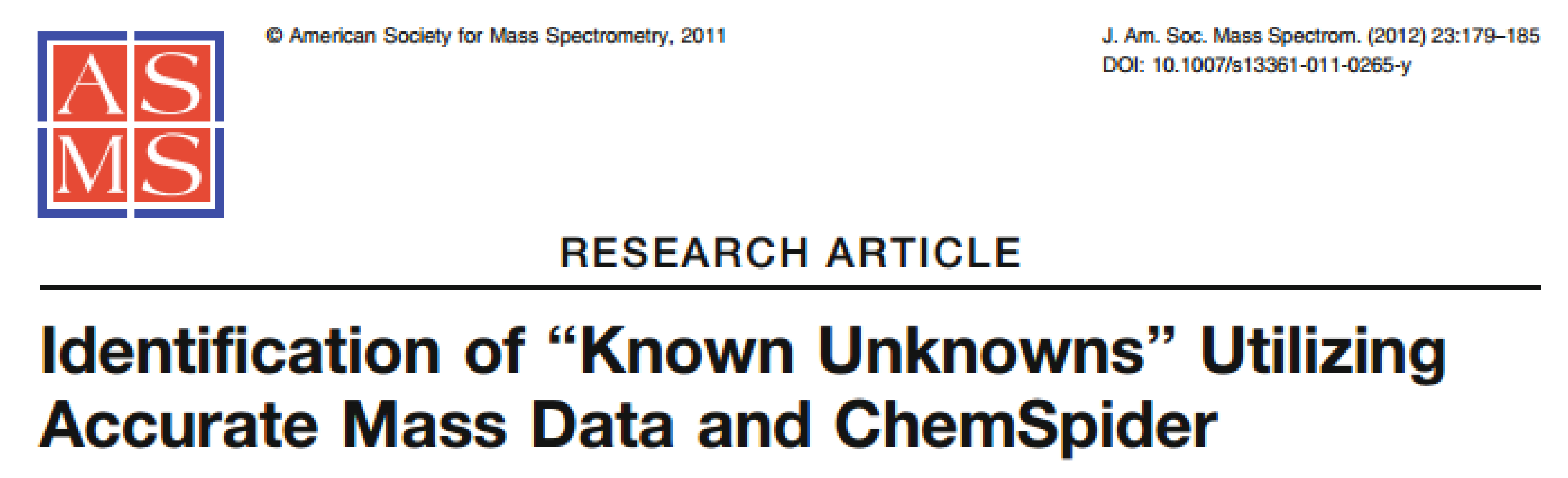
Zincate(2-), [[N,N'-1,2-ethanediylbis[N-[(...
DTXSID: DTXSID8028343
PubChem: 5
CPDAT: 10

EDTA, copper salt
DTXSID: DTXSID0034564
PubChem: 8
CPDAT: 10

Calcium disodium ethylenediaminetetra...
DTXSID: DTXSID2036409
PubChem: 42
CPDAT: 29

Tetrapotassium ethylenediaminetetraace...
DTXSID: DTXSID3036442
PubChem: 25
CPDAT: 36

Candidate ranking using metadata



© American Society for Mass Spectrometry, 2011

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

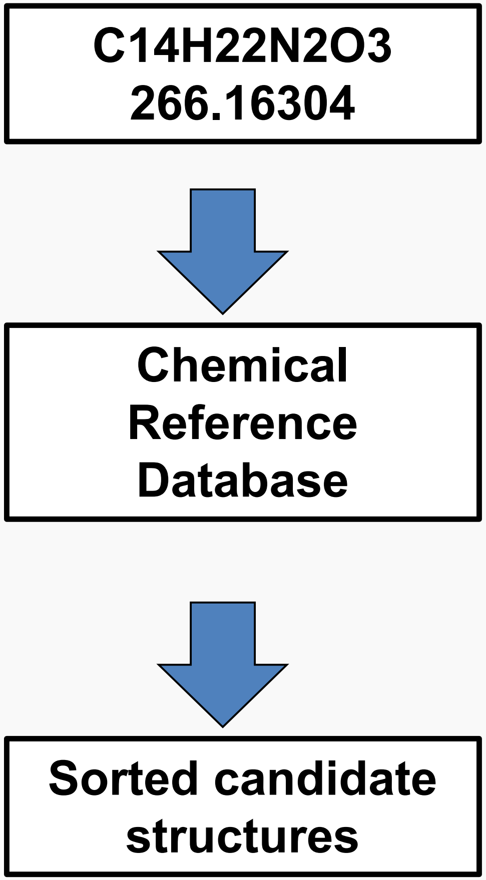
RESEARCH ARTICLE

Identification of “Known Unknowns” Utilizing Accurate Mass Data and ChemSpider

Data Source Ranking of *"known unknowns"*

- A mass and/or formula search is for an ***unknown*** chemical but it is a ***known*** chemical contained within a reference database

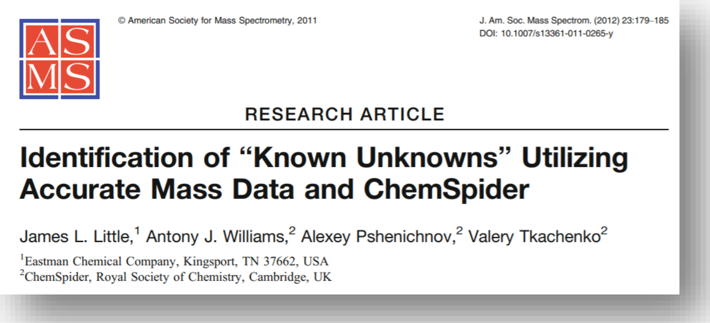
- **Most likely** candidate chemicals have the **most** associated data sources, **most** associated literature articles or both



C14H22N2O3
266.16304

Chemical Reference Database

Sorted candidate structures



© American Society for Mass Spectrometry, 2011

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

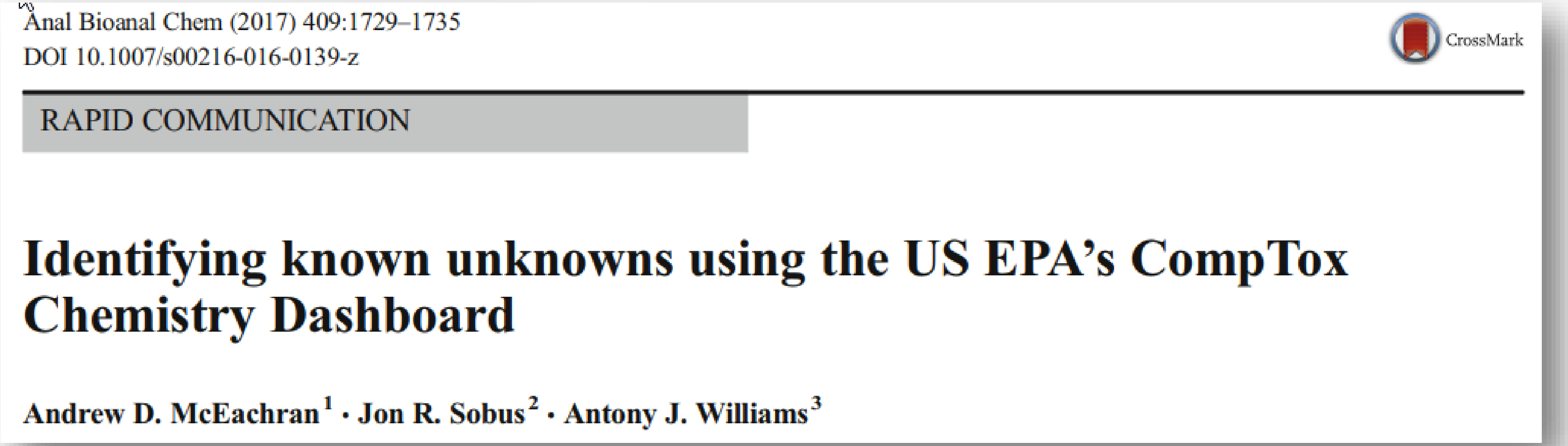
James L. Little,[1] Antony J. Williams,[2] Alexey Pshenichnov,[2] Valery Tkachenko[2]

[1]Eastman Chemical Company, Kingsport, TN 37662, USA
[2]ChemSpider, Royal Society of Chemistry, Cambridge, UK

Dashboard Metadata for Ranking

- Chosen dashboard metadata to rank candidates
  - Associated data sources
    - Lists in the underlying database (more about lists later)
    - Associated data sources in PubChem
    - Specific source types (e.g. water, surfactants, pesticides)
  - Number of associated literature articles (Pubmed)
  - **Chemicals in the environment** – the number of products/categories containing the chemical is an important source of data (from CPDat database)

Comparing Search Performance



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

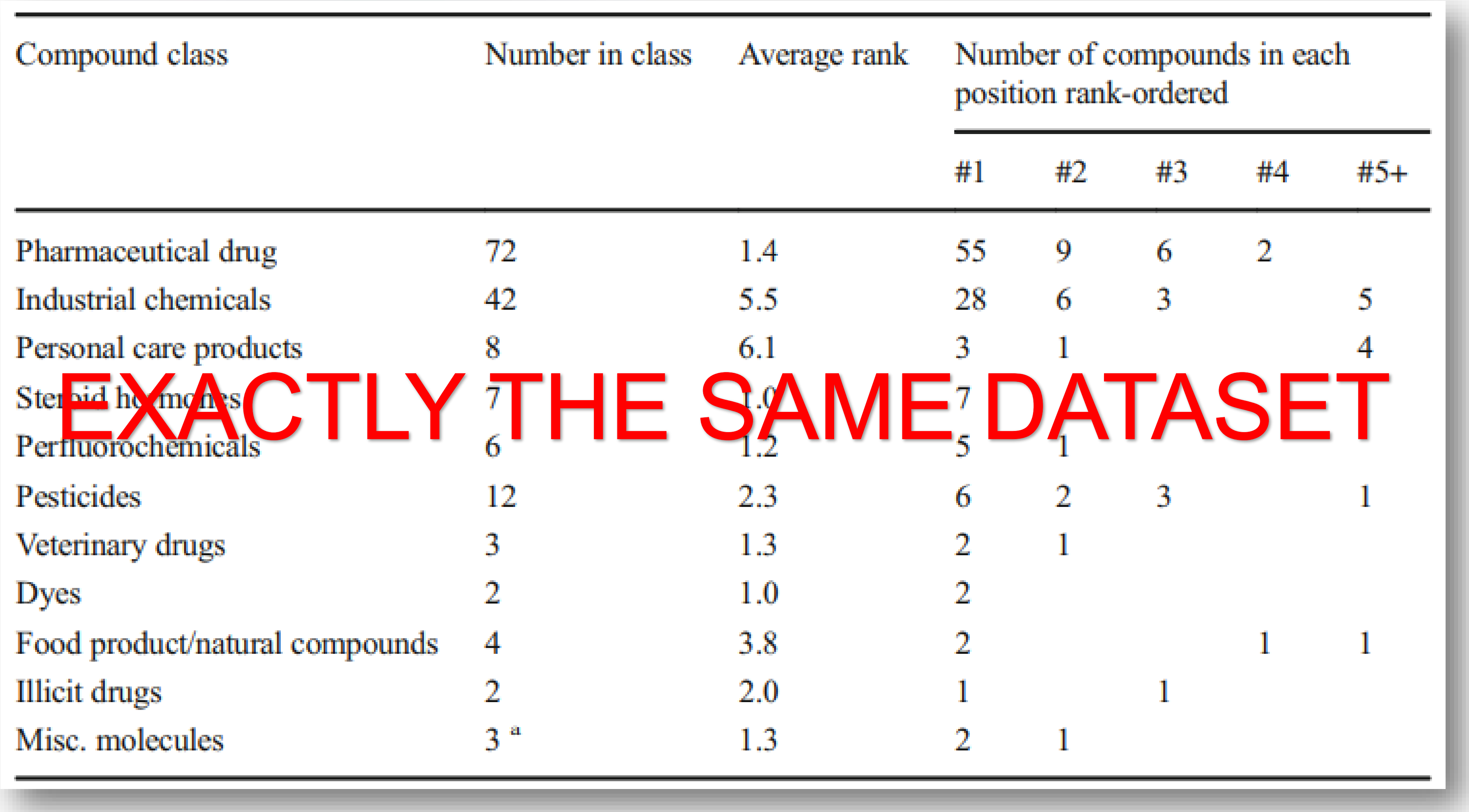
RAPID COMMUNICATION

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

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

- When dashboard contained 720k chemicals
- Only **3%** of ChemSpider size
- What was the comparison in performance?

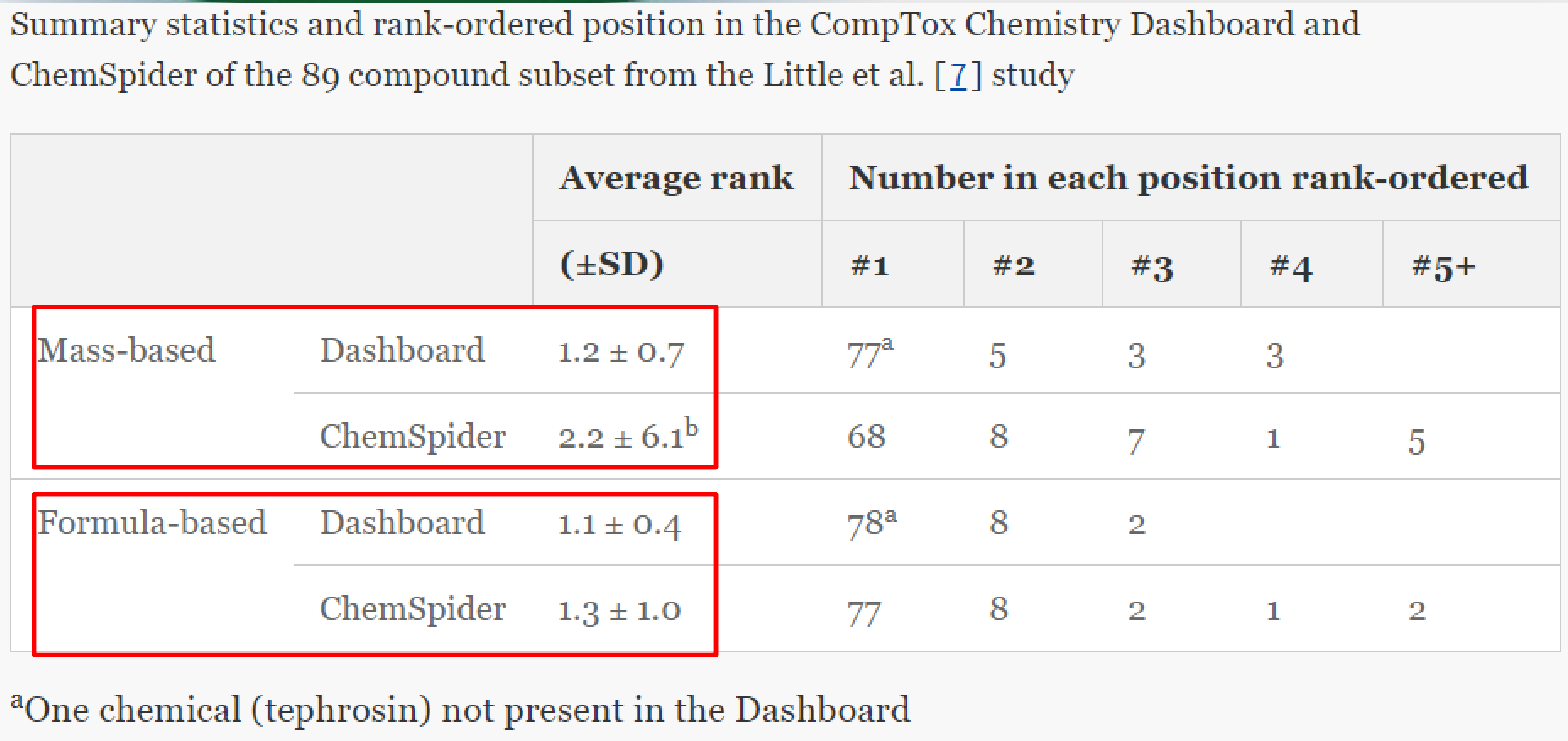
SAME dataset for comparison



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			

EXACTLY THE SAME DATASET

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

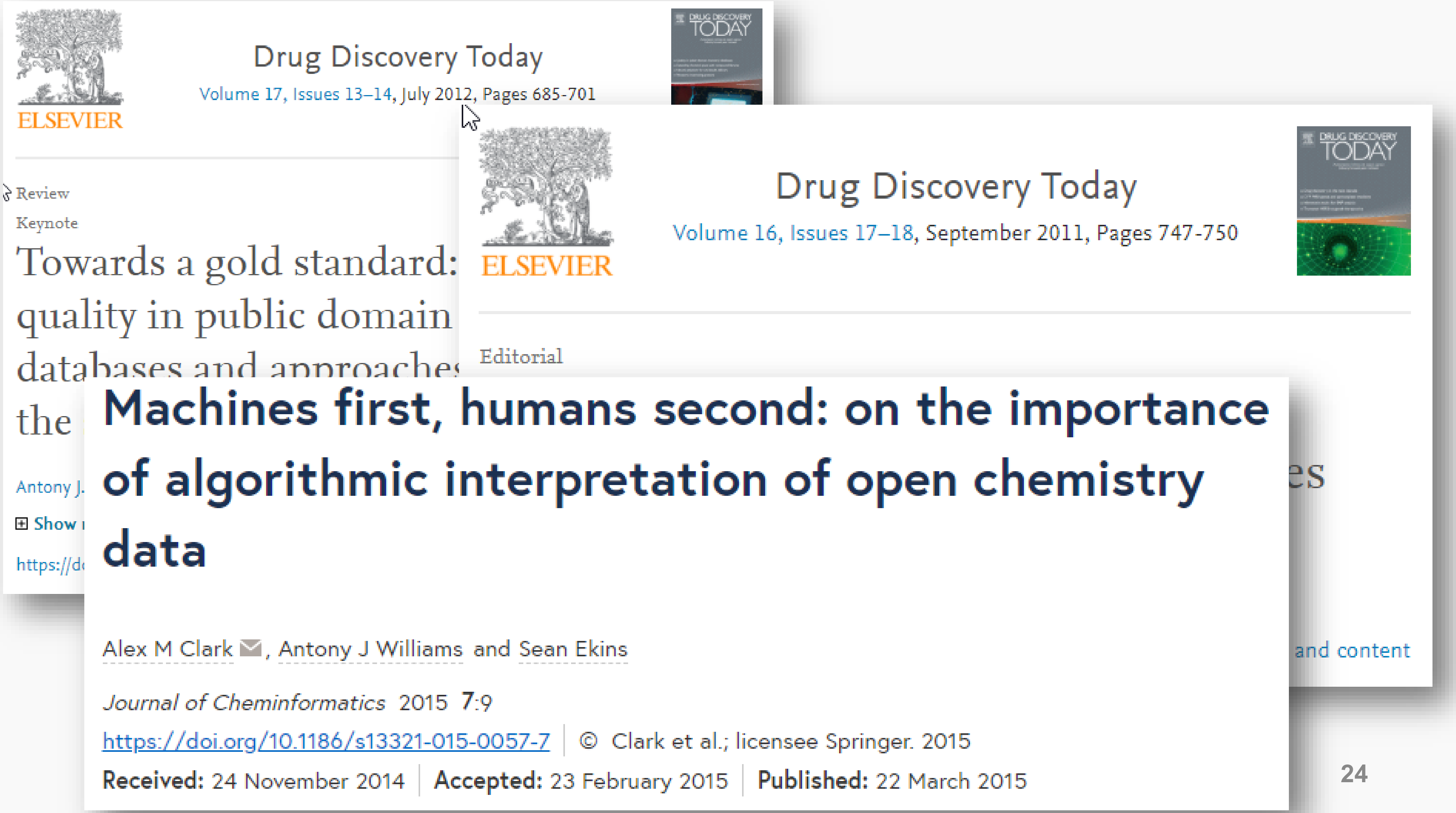
		Average rank	Number in each position rank-ordered				
		(±SD)	#1	#2	#3	#4	#5+
Mass-based	Dashboard	1.2 ± 0.7	77[a]	5	3	3	
	ChemSpider	2.2 ± 6.1[b]	68	8	7	1	5
Formula-based	Dashboard	1.1 ± 0.4	78[a]	8	2		
	ChemSpider	1.3 ± 1.0	77	8	2	1	2

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

For the same 162 chemicals, Dashboard outperforms ChemSpider for both Mass and Formula Ranking

Data Quality is important

- Data quality in free web-based databases!



ELSEVIER

Drug Discovery Today

Volume 17, Issues 13–14, July 2012, Pages 685-701

Review

Keynote

Towards a gold standard: quality in public domain databases and approaches the

Antony J.

⊞ Show

https://d

ELSEVIER

Drug Discovery Today

Volume 16, Issues 17–18, September 2011, Pages 747-750

Editorial

es

and content

Machines first, humans second: on the importance of algorithmic interpretation of open chemistry data

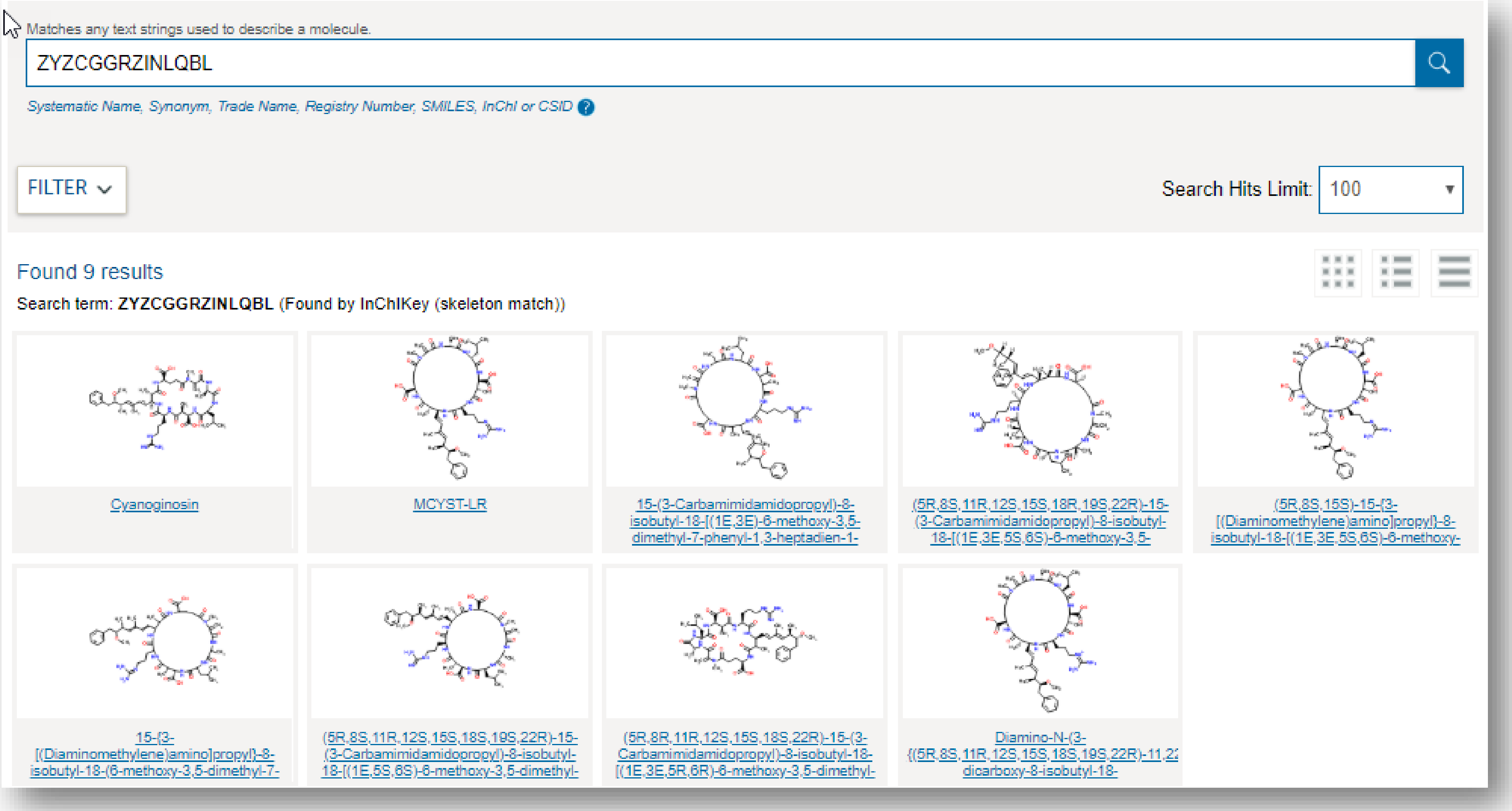
Alex M Clark, Antony J Williams and Sean Ekins

Journal of Cheminformatics 2015 7:9

https://doi.org/10.1186/s13321-015-0057-7 | © Clark et al.; licensee Springer. 2015

Received: 24 November 2014 | **Accepted:** 23 February 2015 | **Published:** 22 March 2015

Will the correct Microcystin LR Stand Up? ChemSpider Skeleton Search



Matches any text strings used to describe a molecule.

ZYZCGGRZINLQBL

Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID

Search Hits Limit: 100

Found 9 results

Search term: **ZYZCGGRZINLQBL** (Found by InChIKey (skeleton match))

Cyanoginosin

MCYST-LR

15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-

(5R,8S,11R,12S,15S,18R,19S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E,5S,6S)-6-methoxy-3,5-

(5R,8S,15S)-15-{3-[(Diaminomethylene)amino]propyl}-8-isobutyl-18-[(1E,3E,5S,6S)-6-methoxy-

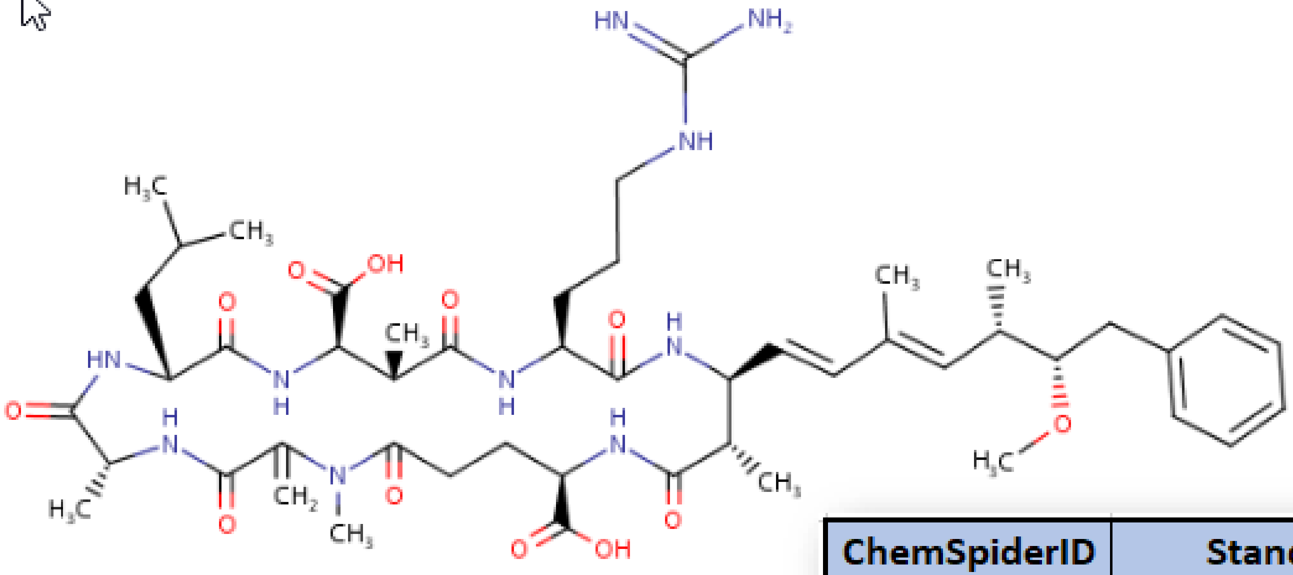
15-{3-[(Diaminomethylene)amino]propyl}-8-isobutyl-18-(6-methoxy-3,5-dimethyl-7-

(5R,8S,11R,12S,15S,18S,19S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,5S,6S)-6-methoxy-3,5-dimethyl-

(5R,8R,11R,12S,15S,18S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E,5R,6R)-6-methoxy-3,5-dimethyl-

Diamino-N-(3-{(5R,8S,11R,12S,15S,18S,19S,22R)-11,2? dicarboxy-8-isobutyl-18-

Comparing ChemSpider Structures



ChemSpiderID	Standard InChIKey Stereolayer
WIKIPEDIA	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
CompTox	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
4941647	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
393078	t28-,29-,30-,31+,34-,35-,36+,**37-**,38-,40+
57618348	t28-,29-,30-,31+,34-,35-,36+,**37-**,38-,40+
29342071	t28-,29-,30-,31+,**34+**,35-,36+,**37-**,38-,40+
7987594	t28-,**29?**,**30?**,31+,**34?**,35-,**36?**,**37-**,38-,**40?**
22900854	t28-,**29?**,**30+**,**31-**,**34+**,**35+**,**36-**,**37-**,38-,**40-**
19692240	NONE
2831283	NONE

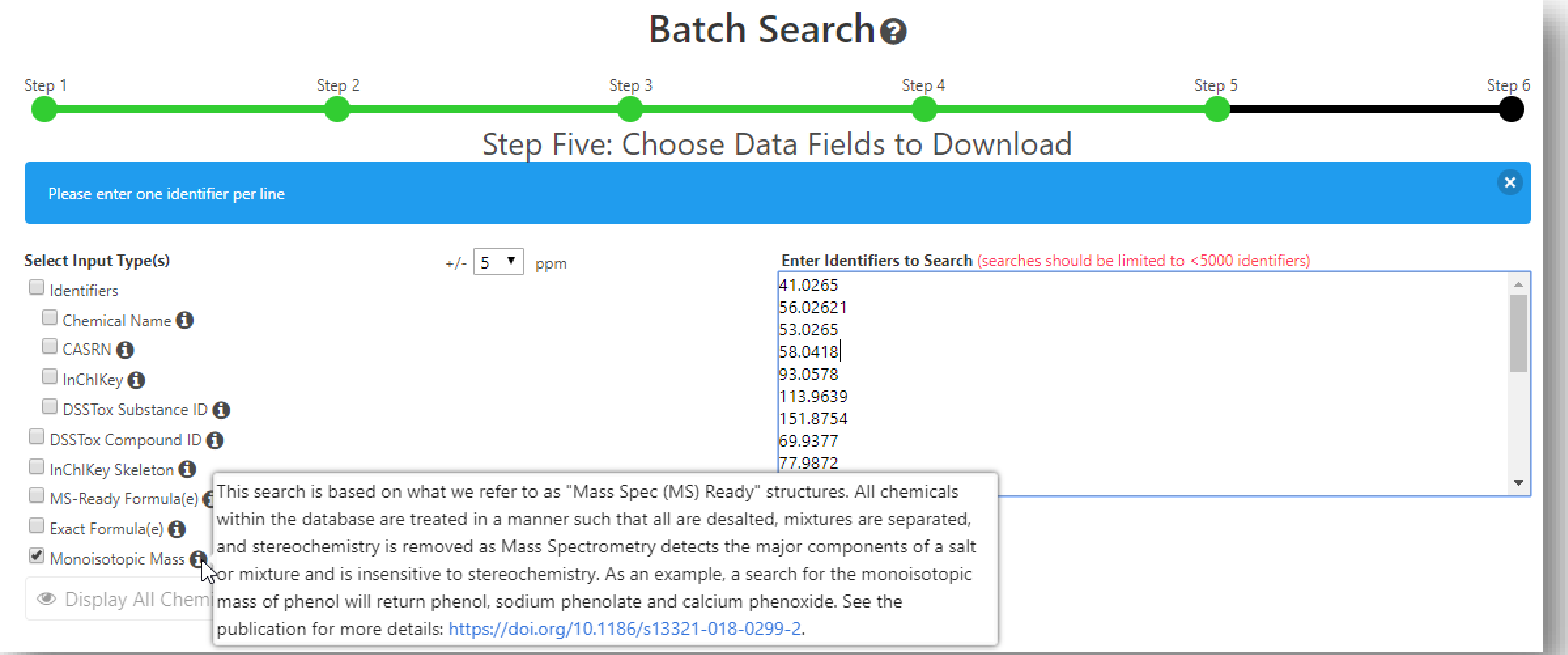
Batch Searching mass and formula

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



Batch Search

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Step Five: Choose Data Fields to Download

Please enter one identifier per line

Select Input Type(s)

- ☐ Identifiers
  - ☐ Chemical Name
  - ☐ CASRN
  - ☐ InChIKey
  - ☐ DSSTox Substance ID
- ☐ DSSTox Compound ID
- ☐ InChIKey Skeleton
- ☐ MS-Ready Formula(e)
- ☐ Exact Formula(e)
- ☑ Monoisotopic Mass

+/- 5 ppm

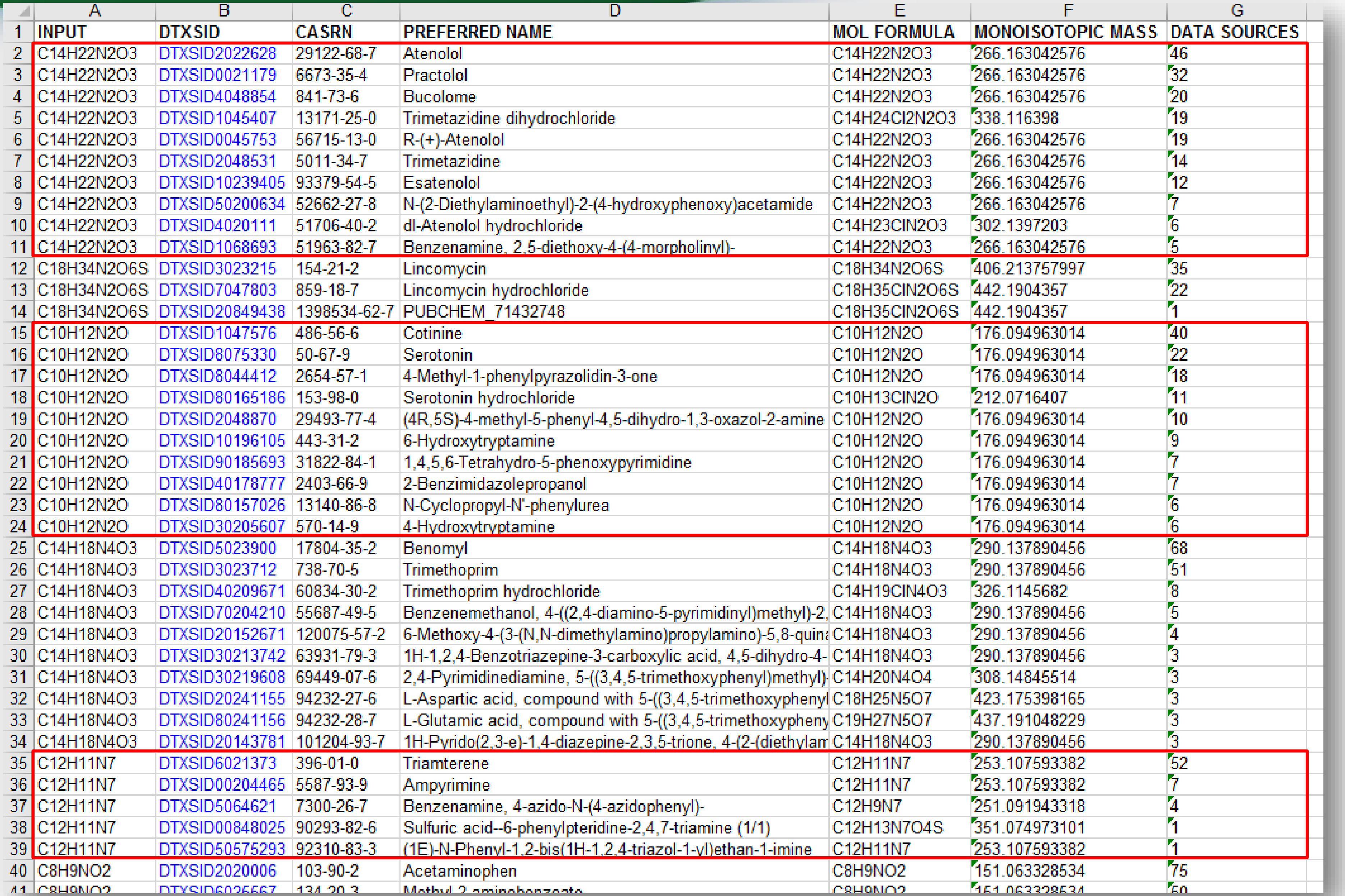
Enter Identifiers to Search (searches should be limited to <5000 identifiers)

```
41.0265
56.02621
53.0265
58.0418
93.0578
113.9639
151.8754
69.9377
77.9872
```

This search is based on what we refer to as "Mass Spec (MS) Ready" structures. All chemicals within the database are treated in a manner such that all are desalted, mixtures are separated, and stereochemistry is removed as Mass Spectrometry detects the major components of a salt or mixture and is insensitive to stereochemistry. As an example, a search for the monoisotopic mass of phenol will return phenol, sodium phenolate and calcium phenoxide. See the publication for more details: https://doi.org/10.1186/s13321-018-0299-2.

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

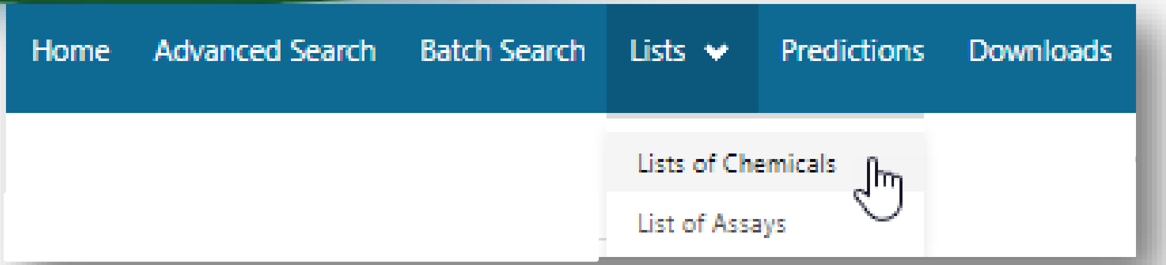
EPA United States Environmental Protection Agency

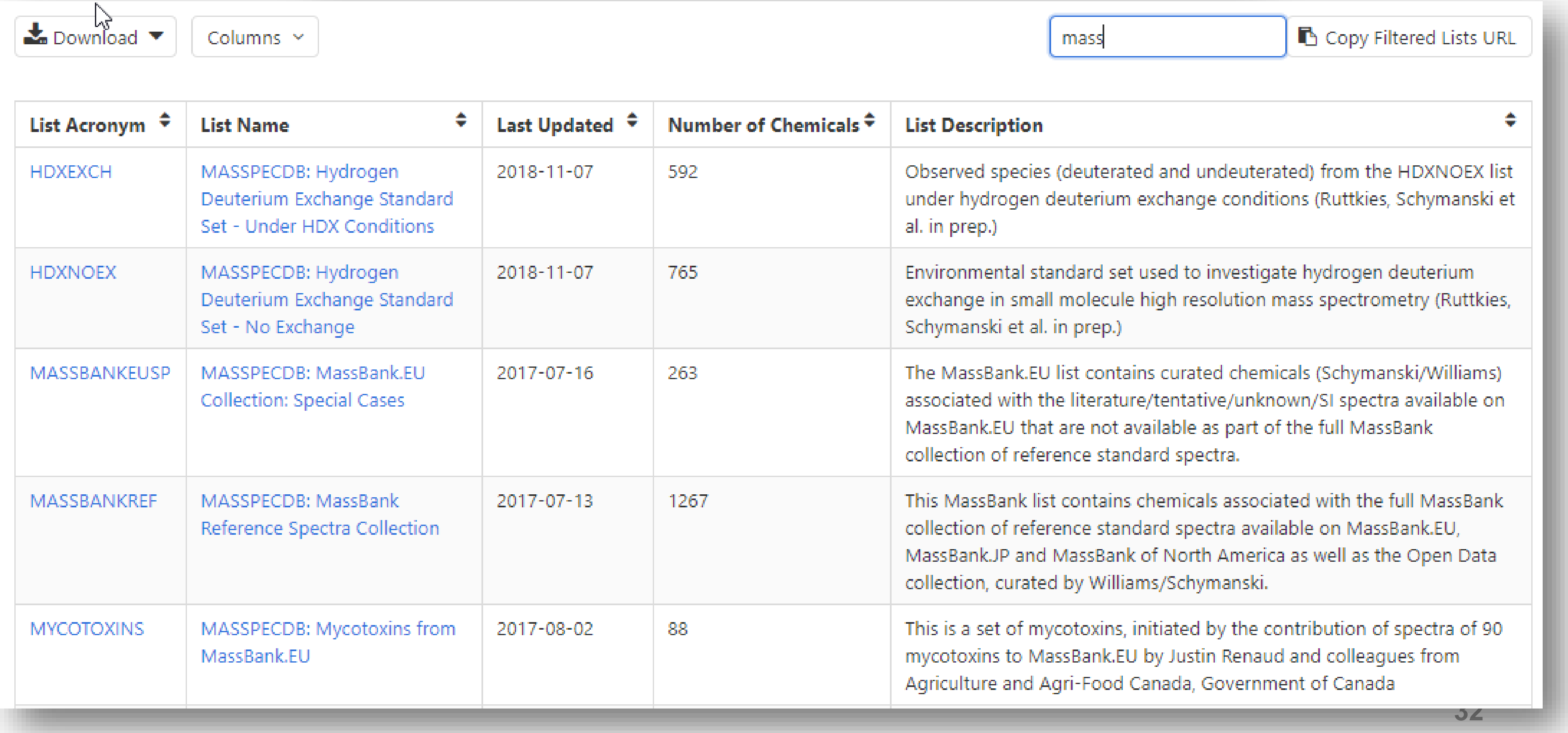


	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	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-quina	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	50

Chemical Lists

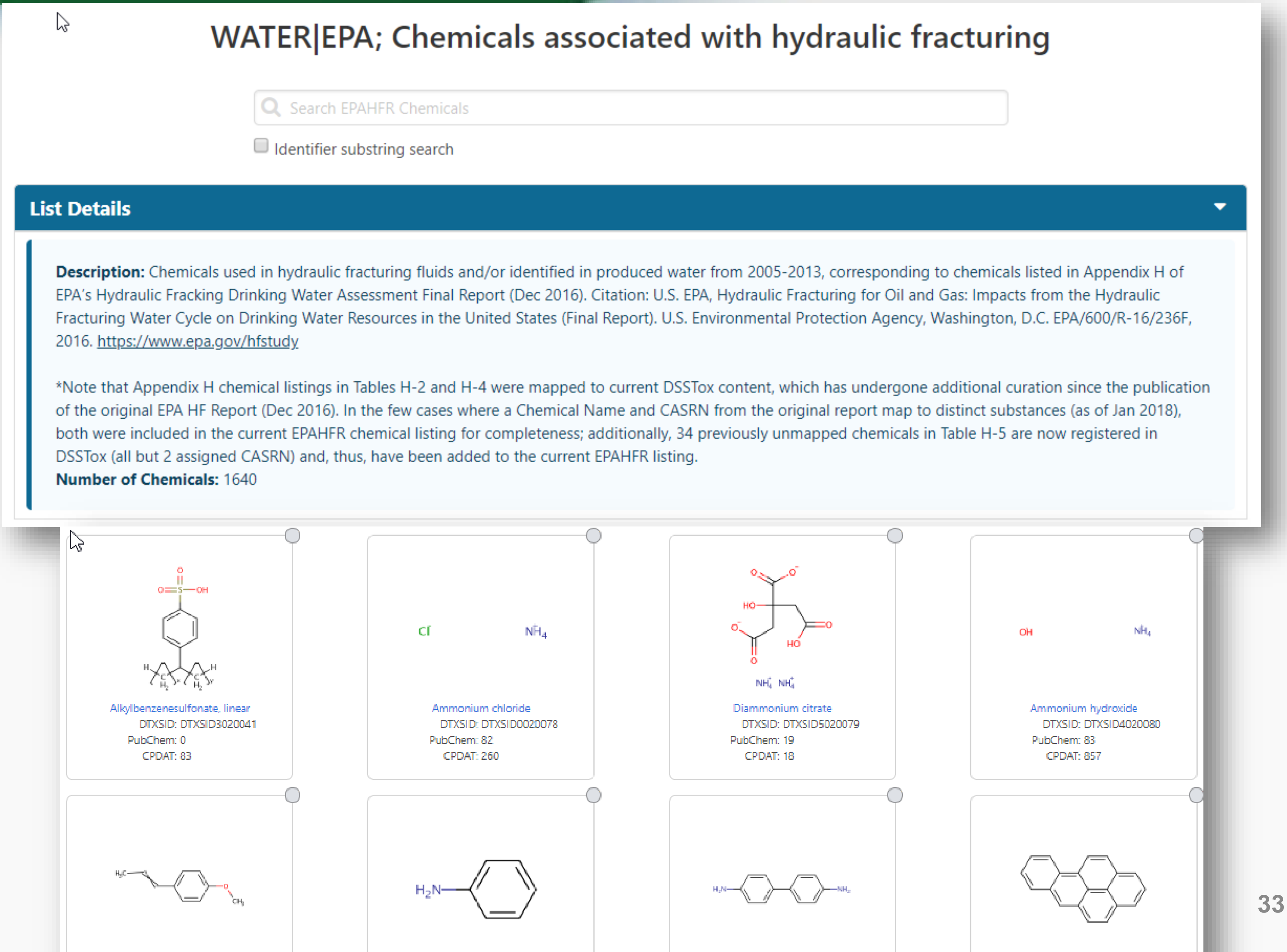
Chemical Lists





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

EPAHFR: Hydraulic Fracturing



WATER|EPA; Chemicals associated with hydraulic fracturing

Search EPAHFR Chemicals

□ Identifier substring search

List Details

Description: Chemicals used in hydraulic fracturing fluids and/or identified in produced water from 2005-2013, corresponding to chemicals listed in Appendix H of EPA's Hydraulic Fracking Drinking Water Assessment Final Report (Dec 2016). Citation: U.S. EPA, Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). U.S. Environmental Protection Agency, Washington, D.C. EPA/600/R-16/236F, 2016. https://www.epa.gov/hfstudy

*Note that Appendix H chemical listings in Tables H-2 and H-4 were mapped to current DSSTox content, which has undergone additional curation since the publication of the original EPA HF Report (Dec 2016). In the few cases where a Chemical Name and CASRN from the original report map to distinct substances (as of Jan 2018), both were included in the current EPAHFR chemical listing for completeness; additionally, 34 previously unmapped chemicals in Table H-5 are now registered in DSSTox (all but 2 assigned CASRN) and, thus, have been added to the current EPAHFR listing.

Number of Chemicals: 1640

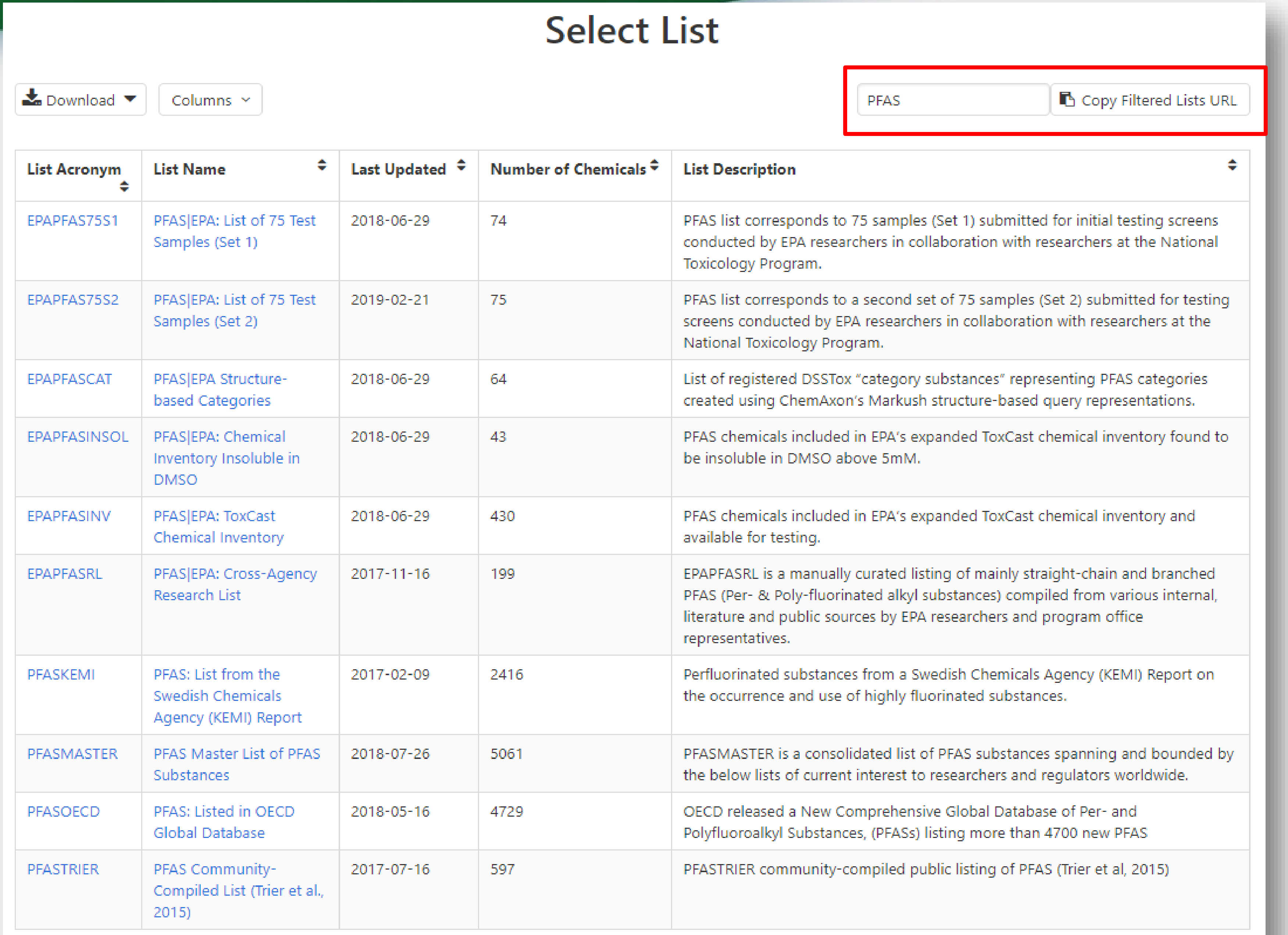
Alkylbenzenesulfonate, linear
DTXSID: DTXSID3020041
PubChem: 0
CPDAT: 83

Ammonium chloride
DTXSID: DTXSID0020078
PubChem: 82
CPDAT: 260

Diammonium citrate
DTXSID: DTXSID5020079
PubChem: 19
CPDAT: 18

Ammonium hydroxide
DTXSID: DTXSID4020080
PubChem: 83
CPDAT: 857

PFAS lists of Chemicals



Select List

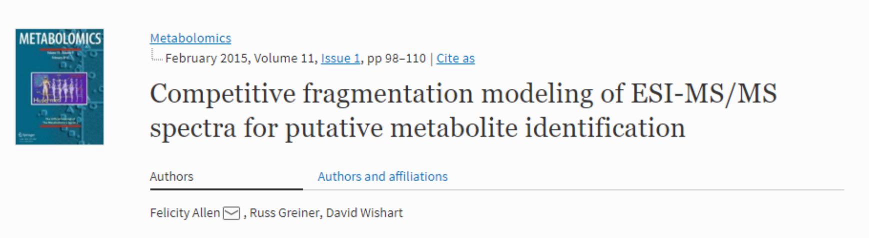
List Acronym	List Name	Last Updated	Number of Chemicals	List Description
EPAPFAS75S1	PFAS|EPA: List of 75 Test Samples (Set 1)	2018-06-29	74	PFAS list corresponds to 75 samples (Set 1) submitted for initial testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFAS75S2	PFAS|EPA: List of 75 Test Samples (Set 2)	2019-02-21	75	PFAS list corresponds to a second set of 75 samples (Set 2) submitted for testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFASCAT	PFAS|EPA Structure-based Categories	2018-06-29	64	List of registered DSSTox "category substances" representing PFAS categories created using ChemAxon's Markush structure-based query representations.
EPAPFASINSOL	PFAS|EPA: Chemical Inventory Insoluble in DMSO	2018-06-29	43	PFAS chemicals included in EPA's expanded ToxCast chemical inventory found to be insoluble in DMSO above 5mM.
EPAPFASINV	PFAS|EPA: ToxCast Chemical Inventory	2018-06-29	430	PFAS chemicals included in EPA's expanded ToxCast chemical inventory and available for testing.
EPAPFASRL	PFAS|EPA: Cross-Agency Research List	2017-11-16	199	EPAPFASRL is a manually curated listing of mainly straight-chain and branched PFAS (Per- & Poly-fluorinated alkyl substances) compiled from various internal, literature and public sources by EPA researchers and program office representatives.
PFASKEMI	PFAS: List from the Swedish Chemicals Agency (KEMI) Report	2017-02-09	2416	Perfluorinated substances from a Swedish Chemicals Agency (KEMI) Report on the occurrence and use of highly fluorinated substances.
PFASMASTER	PFAS Master List of PFAS Substances	2018-07-26	5061	PFASMASTER is a consolidated list of PFAS substances spanning and bounded by the below lists of current interest to researchers and regulators worldwide.
PFASOECD	PFAS: Listed in OECD Global Database	2018-05-16	4729	OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs) listing more than 4700 new PFAS
PFASTRIER	PFAS Community-Compiled List (Trier et al., 2015)	2017-07-16	597	PFASTRIER community-compiled public listing of PFAS (Trier et al, 2015)

Research in Progress

Predicted Mass Spectra

http://cfmid.wishartlab.com/

CFM-ID
Competitive Fragmentation Modeling for Metabolite Identification



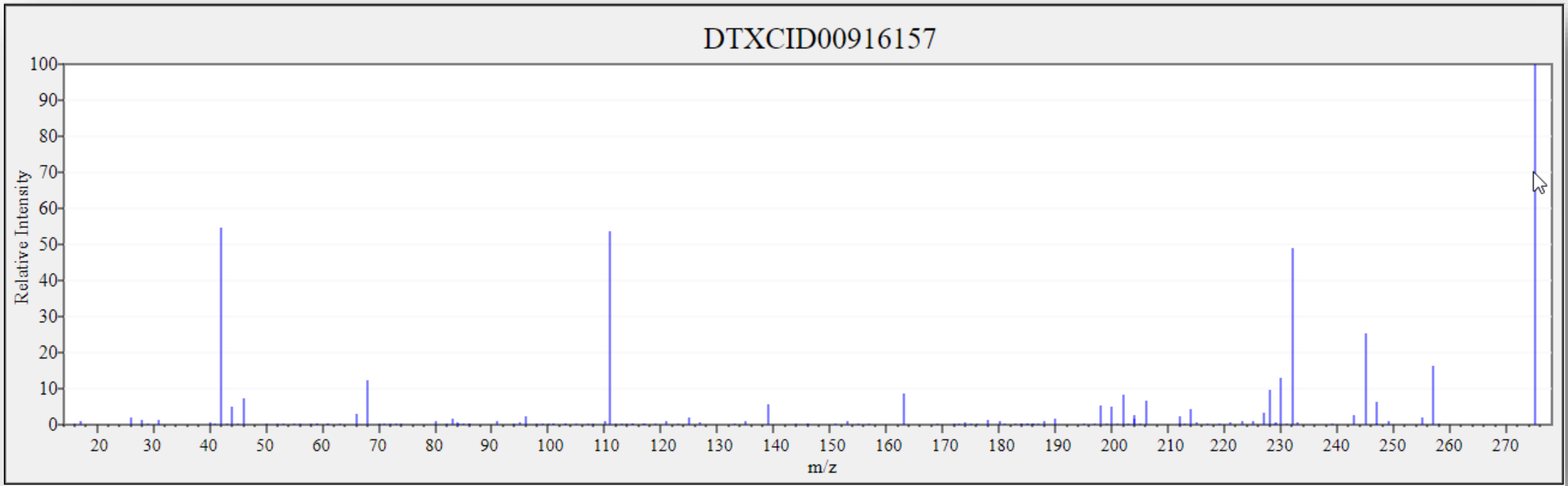
Metabolomics
February 2015, Volume 11, Issue 1, pp 98–110 | Cite as

Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification

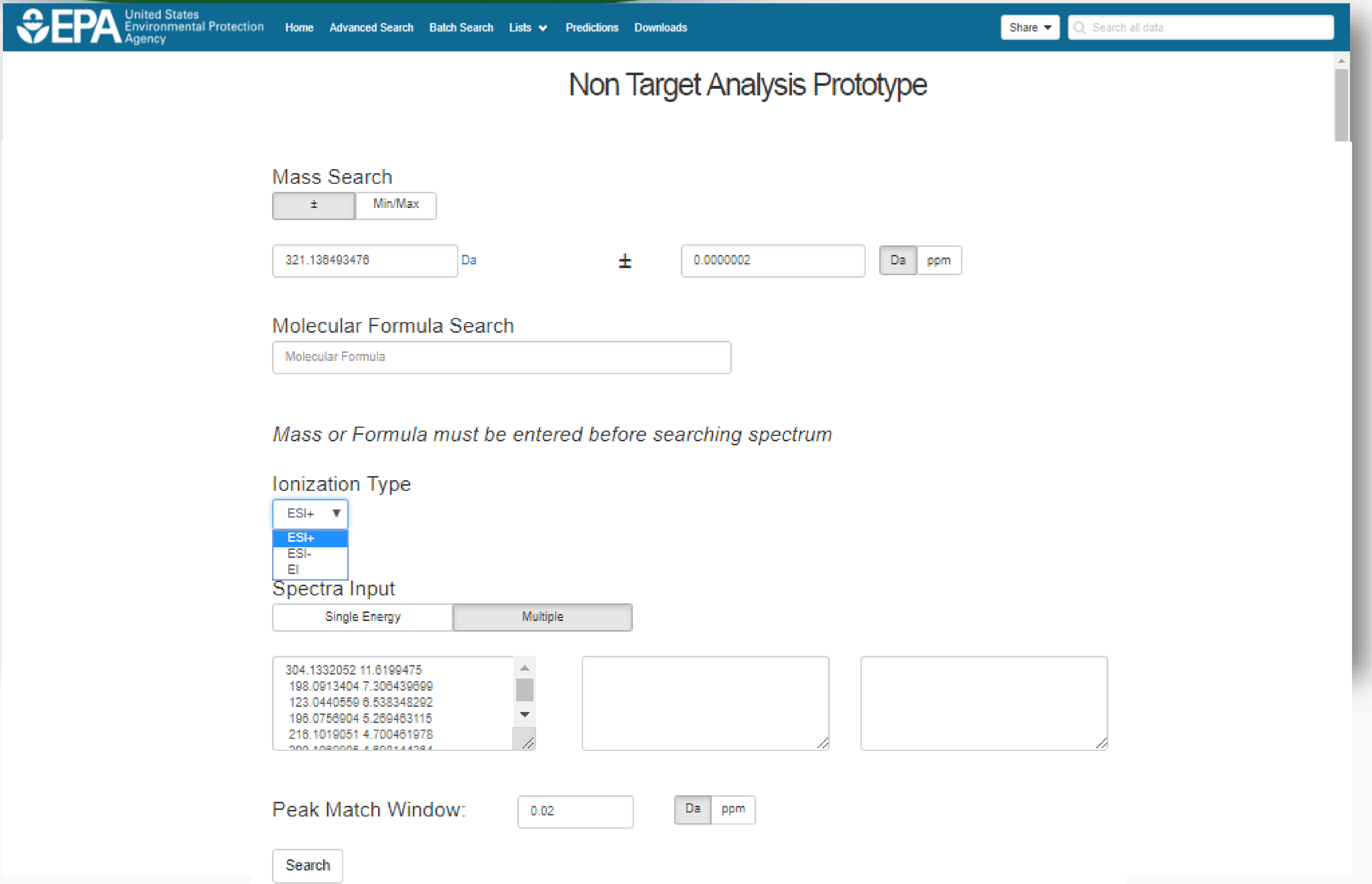
Authors | Authors and affiliations

Felicity Allen, Russ Greiner, David Wishart

- 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



Non Target Analysis Prototype

Mass Search

321.136493476 Da ± 0.0000002 Da ppm

Molecular Formula Search

Mass or Formula must be entered before searching spectrum

Ionization Type

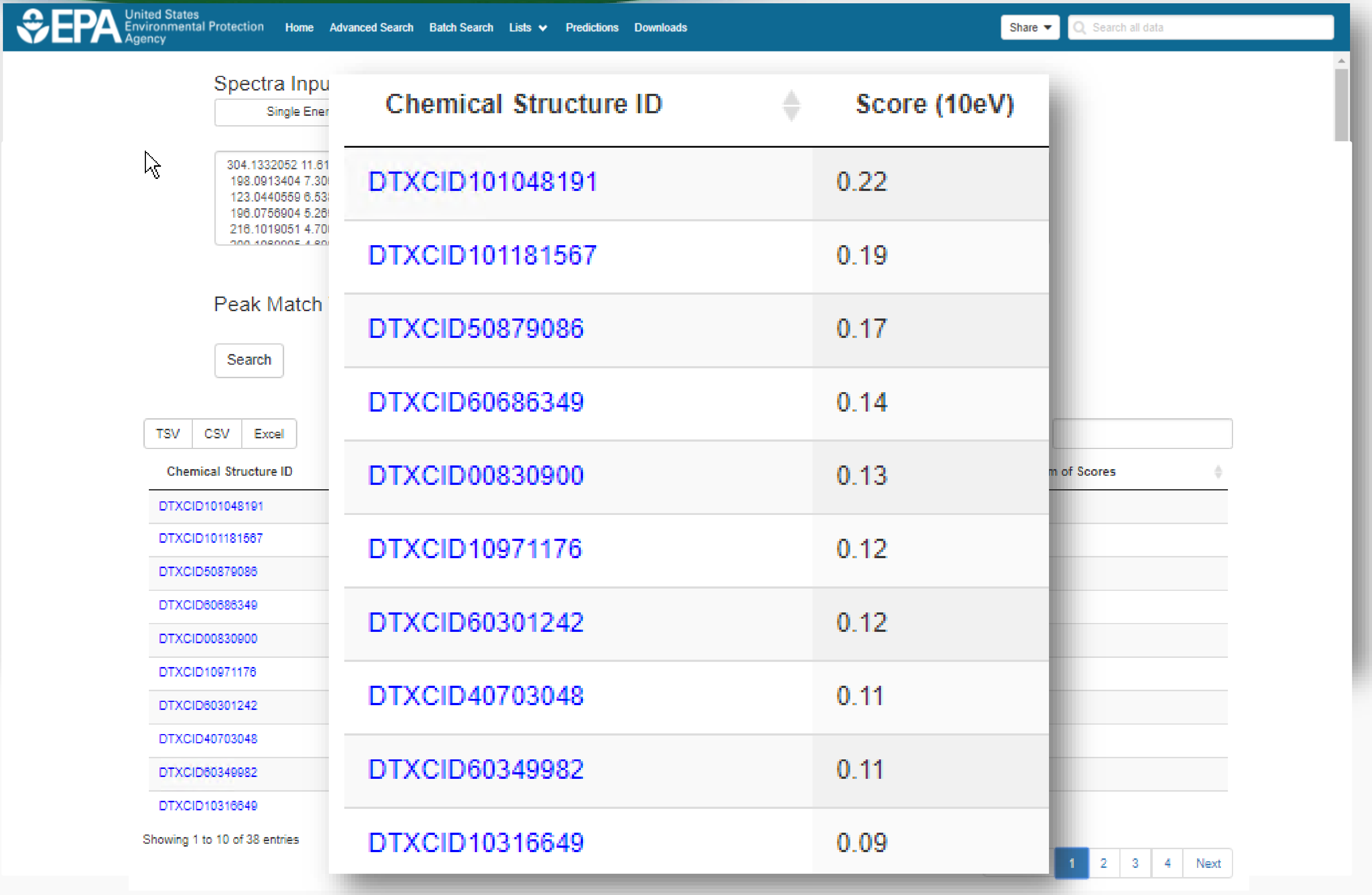
ESI+

Spectra Input

```
304.1332052 11.6199475
 198.0913404 7.306439699
 123.0440559 6.538348292
 196.0756904 5.269463115
 216.1019051 4.700461978
```

Peak Match Window: 0.02 Da ppm

Search Expt. vs. Predicted Spectra

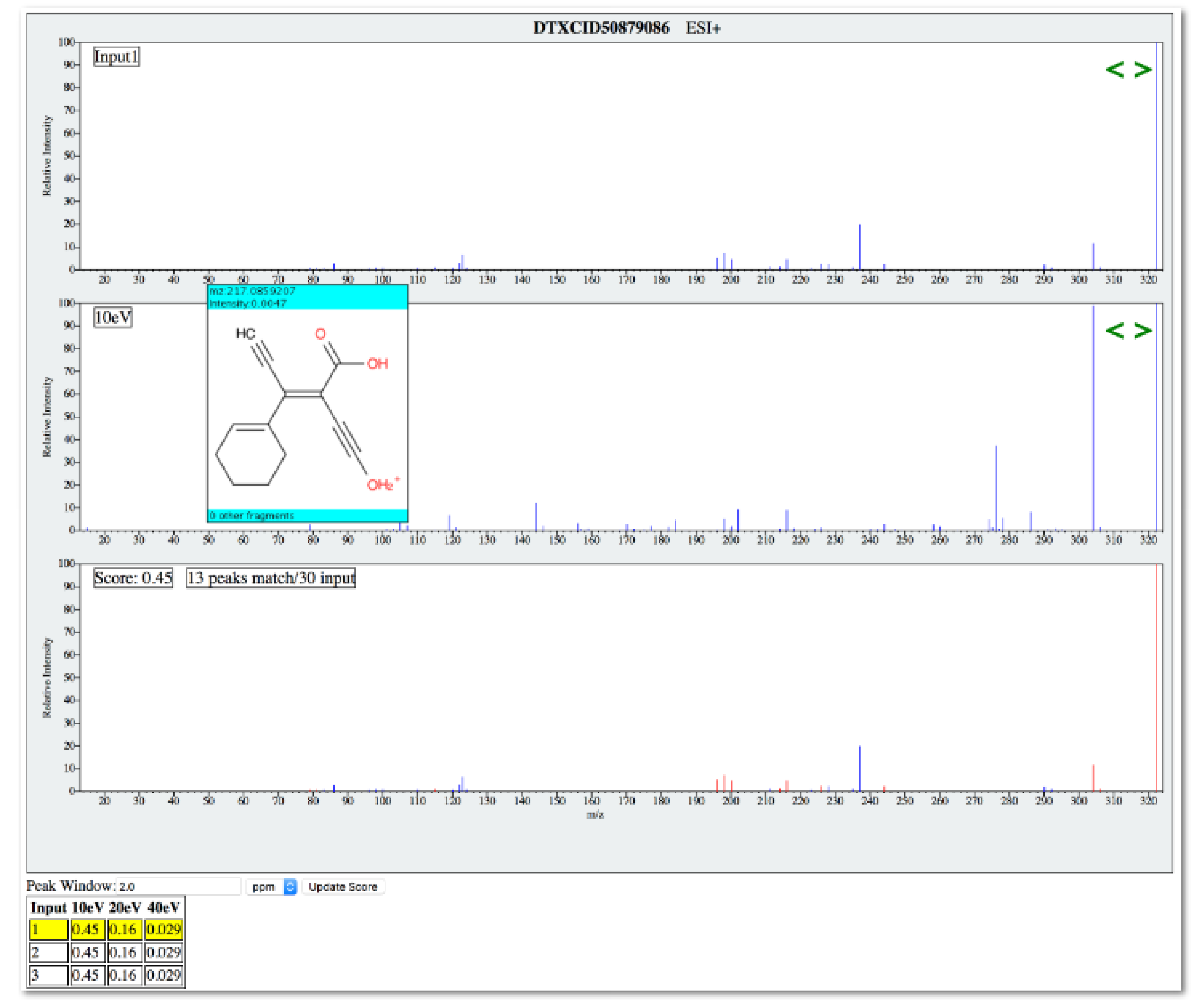


```
304.1332052 11.61
198.0913404 7.30
123.0440559 6.53
196.0756904 5.26
216.1019051 4.70
```

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

Showing 1 to 10 of 38 entries

Spectral Viewer Comparison



DTXCID50879086 ESI+

Input1

10eV

mz:217.0859207
Intensity:0.0047

0 other fragments

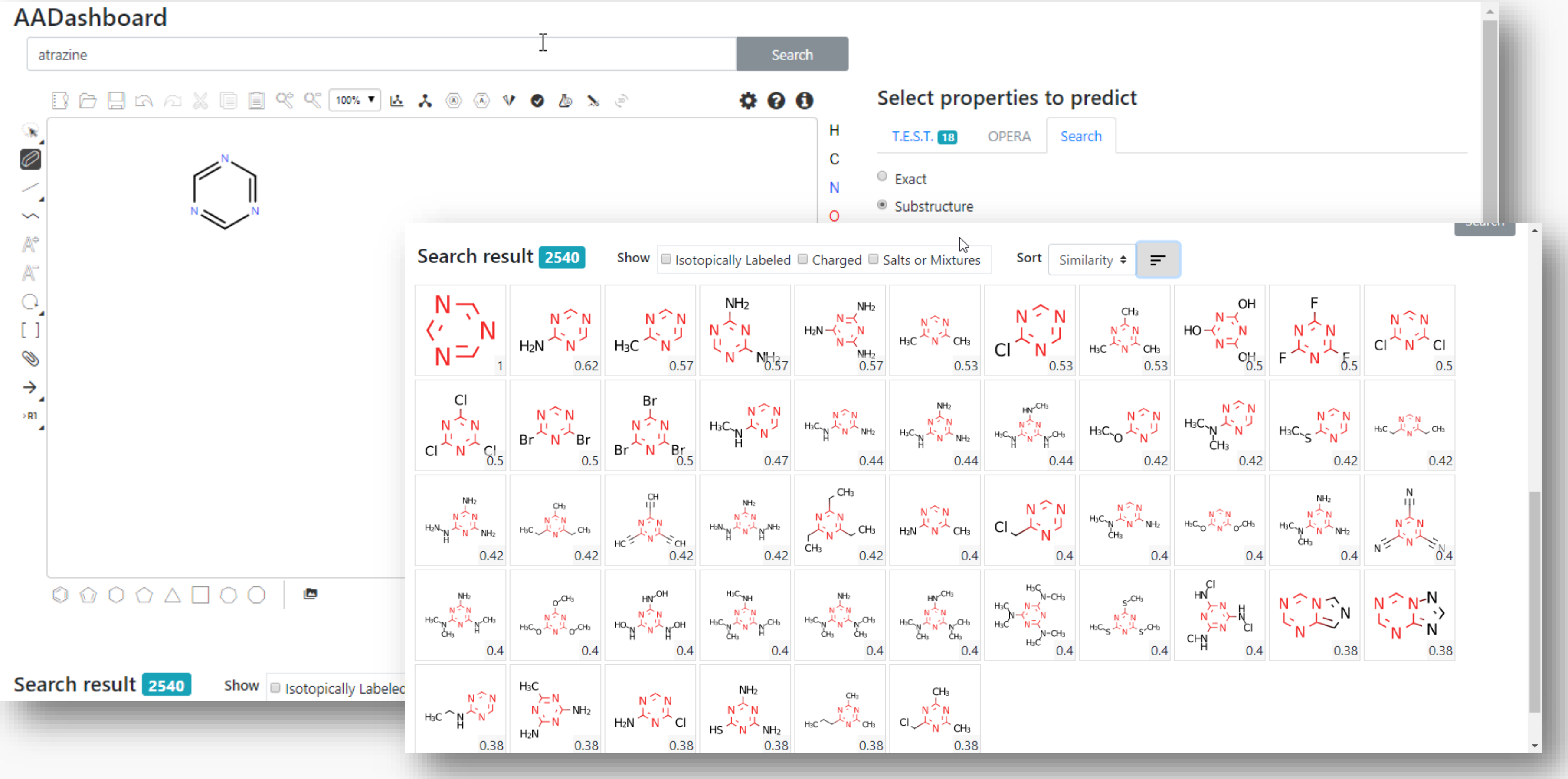
Score: 0.45

13 peaks match/30 input

Peak Window: 2.0 ppm Update Score

Input	10eV	20eV	40eV
1	0.45	0.16	0.029
2	0.45	0.16	0.029
3	0.45	0.16	0.029

Prototype Development



AADashboard

atrazine

Search

Select properties to predict

T.E.S.T. 18 | OPERA | Search

○ Exact

● Substructure

Search result 2540

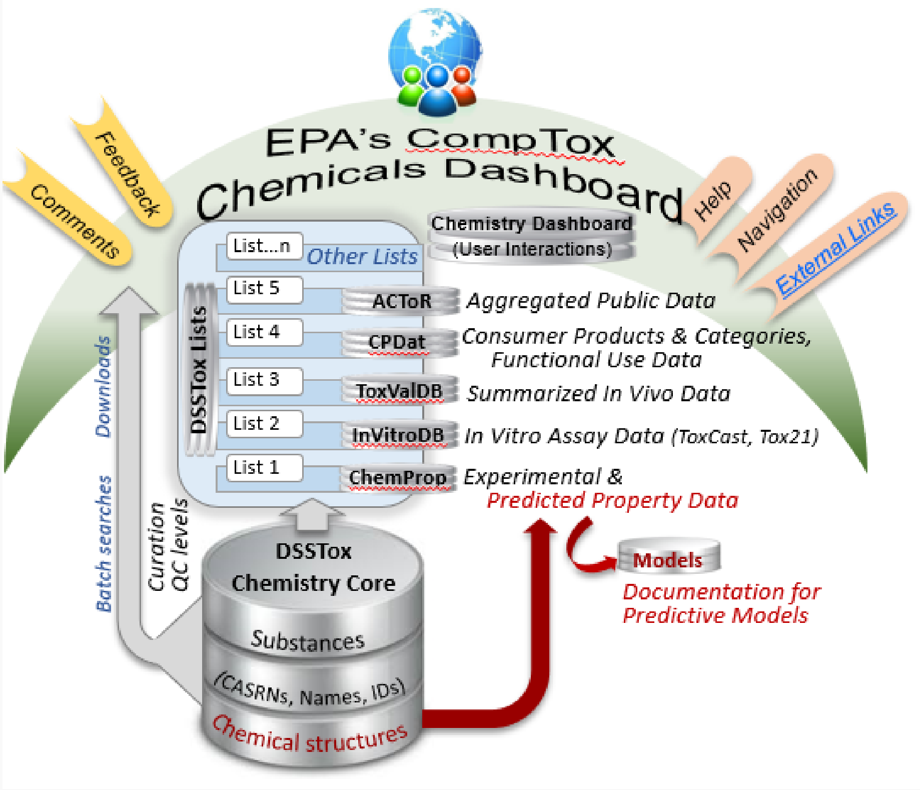
Show □ Isotopically Labeled □ Charged □ Salts or Mixtures

Sort Similarity

Search result 2540 Show □ Isotopically Labeled

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
- New developments in progress, especially API development, will be very enabling…



Acknowledgements

- IT Development team – especially Jeff Edwards and Jeremy Dunne
- Chris Grulke for the ChemReg system
- Andrew McEachran (now at Agilent)
- The curation team focused on data quality

Contact

Antony Williams

US EPA Office of Research and Development

Center for Computational Toxicology and Exposure

EMAIL: Williams.Antony@epa.gov

ORCID: https://orcid.org/0000-0002-2668-4821