



# Leveraging chemistry data to improve exposure analyses using the EPA's CompTox Dashboard

*Andrew D. McEachran*

*Katherine Phillips, Kristin Isaacs, Seth Newton, Chris Grulke, Jon Sobus, and Antony J. Williams*

AGRO 29

*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*

# Exposure Science in the 21<sup>st</sup> Century: What are Key Drivers?

1) Understanding causes  
of disease

2) Ensuring chemical safety  
and human/eco health

*“...70-90% of disease risks are probably due to differences in environments”*

**EPIDEMIOLOGY**

## Environment and Disease Risks

Stephen M. Rappaport and Martyn T. Smith

Although the risks of developing chronic diseases are attributed to both genetic and environmental factors, 70 to 90% of disease risks are probably due to differences in environments (1-3). Yet, epidemiologists increasingly use genome-wide association studies (GWAS) to investigate diseases, while relying on questionnaires to characterize “environmental exposures.” This is because GWAS represent the only approach for exploring the totality of any risk factor (genes, in this case) associated with disease prevalence. Moreover, the value of costly genetic information is diminished when inaccurate and imprecise environmental data lead to biased inferences regarding gene-environment interactions (4). A more comprehensive and quantitative view of environmental exposure is needed if epidemiologists are to discover the major causes of chronic diseases.

An obstacle to identifying the most important environmental exposures is the fragmentation of epidemiological research along lines defined by different factors. When epidemiologists investigate environmental risks, they tend to concentrate on a particular category of exposures involving air and water pollution, occupation, diet and obesity, stress and behavior, or types of infection. This slicing of the disease pie along parochial lines leads to scientific separation and confuses the definition of “environmental exposures.” In fact, all of these exposure categories can contribute to chronic diseases and should be investigated collectively rather than separately.

To develop a more cohesive view of environmental exposure, it is important to recognize that toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the “environment” as the body’s internal chemical environment and “exposures” as the amounts of biologically active chemicals in this internal environment. Under this view, exposures are not restricted to chemicals (toxics) entering the body from air, water, or food, for example, but also include chemicals produced by inflammation, oxidative stress, lipid peroxidation, infections, gut flora, and other natural processes (5, 6) (see the figure). This internal chemical environment continually fluctuates during life due to changes in external and internal sources, aging, infections, life-style, stress, psychosocial factors, and preexisting diseases.

The term “exposome” refers to the totality of environmental exposures from conception onwards, and has been proposed to be a

A new paradigm is needed to assess how a lifetime of exposure to environmental factors affects the risk of developing chronic diseases.

School of Public Health, University of California, Berkeley, CA 94720-7356, USA. E-mail: srappaport@berkeley.edu

460 22 OCTOBER 2010 VOL 330 SCIENCE www.sciencemag.org  
Published by AAAS

8/10/10

GIVE A DOG A PHONE  
Technology for our furry friends

# NewScientist

WEEKLY November 29 - December 6, 2010

We've made  
150,000 new chemicals



We touch them,  
we wear them, we eat them

## But which ones should we worry about?

SPECIAL REPORT, page 34

THE GOOD FIGHT  
Most violence  
is also virtuous

CHAMBER OF SECRETS  
The greatest ever find  
of early human bones

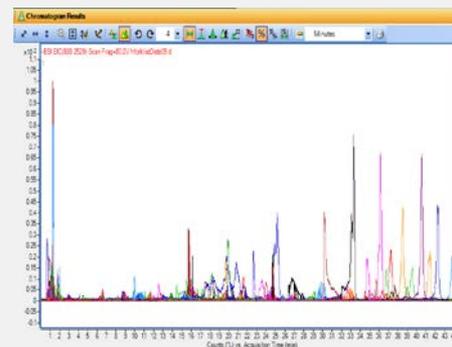
IS IT ALIVE?  
Artificial worm could  
be first digital animal

Science and technology news: www.newscientist.com 130 jobs in science

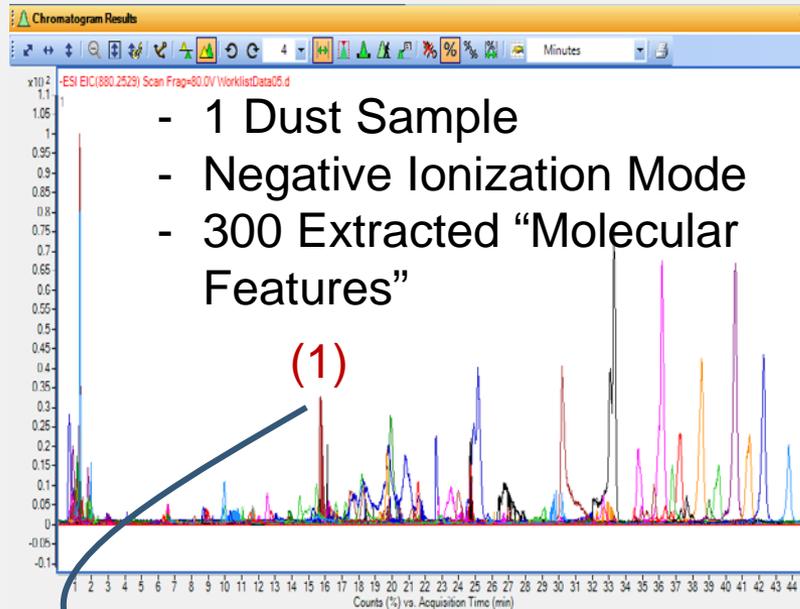
NO2997 US\$5.95 CAN\$ 9.95  
8 74967 304 99 0

# Comparing Analysis Approaches

- Targeted Analysis:
  - We know exactly what we're looking for
  - 10s – 100s of chemicals
  - <<1% of the exposome
- Suspect Screening Analysis (SSA):
  - We have chemicals of interest
  - 100s – 1,000s of chemicals
  - ~5-10% of the exposome
- Non-Targeted Analysis (NTA):
  - We have no preconceived lists
  - 1,000s – 10,000s of chemicals
  - 90-95% of the exposome
  - In dust, soil, food, air, water, products- potential exposure source for plants, animals, and humans



# General Goals of SSA/NTA



1) Prioritize "Molecular Features"

2) Correctly assign formulas

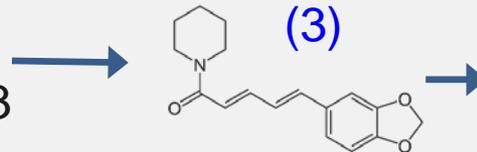
3) Correctly assign structures

4) Identify chemical sources

5) Predict chemical concentrations

EXPOSURE

(2)  
C17H19NO3



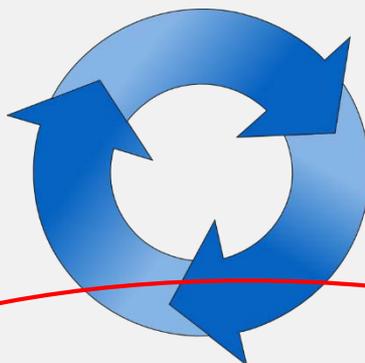
(5)  
12  $\mu\text{g/g}$

# The General Approach

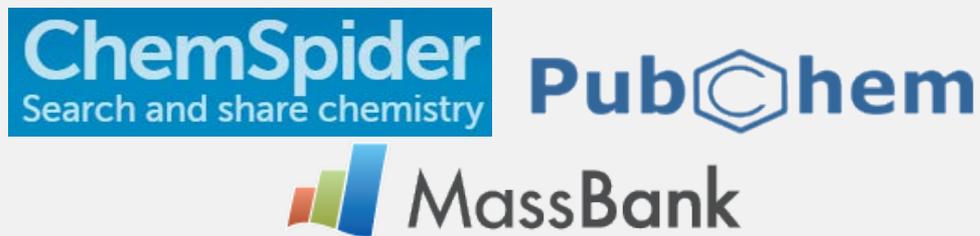
Analytical Instruments



Comp. Tools & Workflows

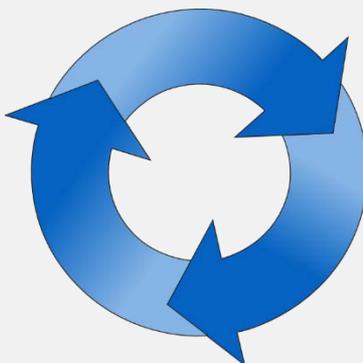


Databases

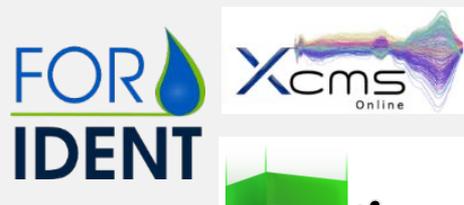


# The General Approach

## Analytical Instruments



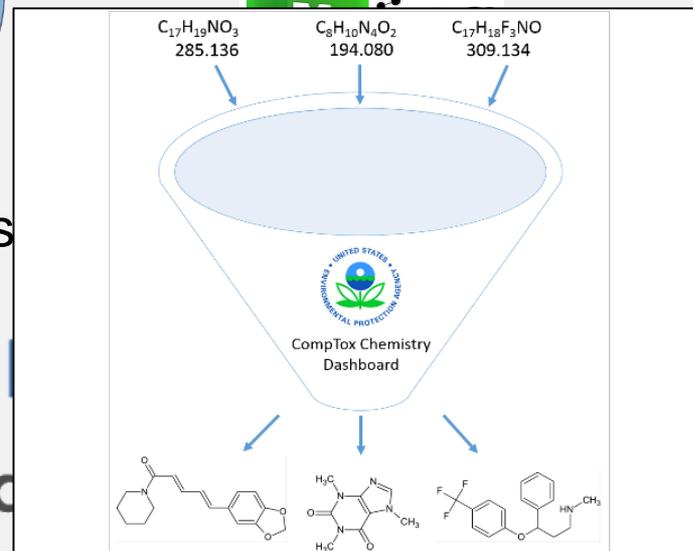
## Comp. Tools & Workflows



## Databases



Pu



# CompTox Dashboard

762 Thousand Chemicals



**Chemicals** [Product/Use Categories](#) [Assay/Gene](#)

Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey

Identifier substring search

See what people are saying, read the dashboard comments!  
Cite the Dashboard Publication [click here](#)

## Latest News

[Read more news](#)

### An article regarding an Excel Version of the Abstract Sifter is published.

March 7th, 2018 at 9:21:27 AM

antages The abstract sifter that is integrated into the Dashboard (for example [here for Atrazine](#)) is available as an Excel add-in. Our recent article on the Abstract Sifter for Excel [has been published](#)

<https://comptox.epa.gov>



### Discover.

[About/Disclaimer](#)  
[Accessibility](#)  
[Privacy](#)

### Connect.

[ACToR](#)  
[DSSTox](#)  
[Downloads](#)

### Ask.

[Contact](#)  
[Help](#)

# Nicotine

54-11-5 | DTXSID1020930

Searched by Approved Name.

## DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

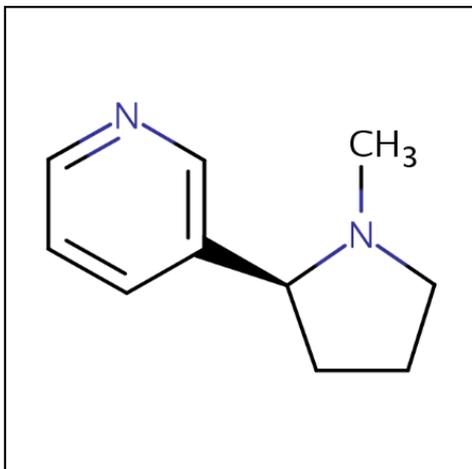
RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

COMMENTS



## Wikipedia

**Nicotine** is a potent parasympathomimetic stimulant and an alkaloid found in the nightshade family of plants. Nicotine acts as an agonist at most nicotinic acetylcholine receptors (nAChRs), except at two nicotinic receptor subunits (nAChR $\alpha$ 9 and nAChR $\alpha$ 10) where it acts as a receptor antagonist. Nicotine is found in the leaves of *Nicotiana rustica*, in concentrations of 2–14%; in the tobacco plant, *Nicotiana tabacum*; in *Duboisia hopwoodii*; and in *Asclepias syriaca*

...  
[Read more](#)

## Intrinsic Properties

 Molecular Formula: C<sub>10</sub>H<sub>14</sub>N<sub>2</sub>  Mol File  Find All Chemicals

 Average Mass: 162.236 g/mol  Isotope Mass Distribution

 Monoisotopic Mass: 162.115698 g/mol

## Structural Identifiers

## Linked Substances

## Presence in Lists

## Record Information

## Quality Control Notes

# Nicotine

54-11-5 | DTXSID1020930

Searched by Approved Name.

DETAILS

EXECUTIVE SUMMARY

**PROPERTIES**

ENV. FATE/TRANSPORT

HAZARD

ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

Property

Summary

Download

Columns

## Summary

Search query

Property	Experimental average	Predicted average	Experimental median	Predicted median	Experimental range	Predicted range	Unit
LogP: Octanol-Water	1.17 (1)	0.751		0.821	1.17	3.85e-2 to 1.18	
Melting Point	-79.0 (3)	12.4	-79.0	13.4	-79.0	-34.4 to 57.3	°C
Boiling Point	247 (2)	249	247	248	247	244 to 254	°C
Vapor Pressure	3.80e-2 (1)	1.70e-2		1.76e-2	3.80e-2	2.39e-3 to 3.03e-2	mmHg
Water Solubility	6.16 (1)	3.74		4.51	6.16	8.00e-2 to 6.63	mol/L
Flash Point	-	99.8		99.8	-	97.9 to 102	°C
Surface Tension	-	38.6		38.6	-	37.7 to 39.6	dyn/cm
Index of Refraction	-	1.54			-	1.54	
Molar Refractivity	-	49.3			-	49.3	cm <sup>3</sup>
Polarizability	-	19.5			-	19.5	Å <sup>3</sup>

# Nicotine

54-11-5 | DTXSID1020930

Searched by Approved Name.

## Exposure Predictions (mg/kg-bw/day)

Download

Columns

Search query

Demographic	Median	95th Percentile
Ages 6-11	1.07e-6	9.90e-5
Ages 12-19	7.88e-7	5.53e-5
Ages 20-65	6.60e-7	4.09e-5
Ages 65+	5.27e-7	2.76e-5
BMI > 30	7.95e-7	4.13e-5
BMI < 30	7.93e-7	5.06e-5
Repro. Age Females	7.35e-7	4.93e-5
Females	8.98e-7	5.34e-5
Males	6.51e-7	5.08e-5
Total	7.08e-7	4.37e-5

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

PRODUCT & USE CATEGORIES

CHEMICAL WEIGHT FRACTION

CHEMICAL FUNCTIONAL USE

TOXICS RELEASE INVENTORY

MONITORING DATA

EXPOSURE PREDICTIONS

PRODUCTION VOLUME

## Batch Search



### Step One: Select Input

Please enter one identifier per line



#### Select Input Type(s)

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

Chemical Data

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

Excel ▼

Download

Customize Results

- Select All
- Select All in Lists

Chemical Identifiers

- DTXSID ⓘ
- Chemical Name ⓘ
- CAS-RN ⓘ
- InChIKey ⓘ
- IUPAC Name ⓘ

Structures

- Mol File ⓘ
- SMILES ⓘ
- InChI String ⓘ
- MS-Ready SMILES ⓘ
- QSAR-Ready SMILES ⓘ

Intrinsic And Predicted Properties

- Molecular Formula ⓘ
- Average Mass ⓘ
- Monoisotopic Mass ⓘ
- TEST Model Predictions ⓘ
- OPERA Model Predictions ⓘ

Metadata

- Curation Level Details ⓘ
- NHANES/Predicted Exposure ⓘ
- Data Sources ⓘ
- Include ToxVal Data Availability ⓘ
- Assay Hit Count
- Number of PubMed Articles ⓘ

Presence in Lists:

- ICCVAM test method evaluation report: in vitro ocular toxicity test methods
- 40CFR355
- A list of all PBDEs (Polybrominated diphenyl ethers)
- A list of all PCBs (Polychlorinated biphenyls)
- A list of polycyclic aromatic hydrocarbons
- Acute exposure guideline levels
- Algal Toxins
- Androgen Receptor Chemicals
- APCRA Chemicals for Prospective Analysis
- APCRA Chemicals for Retrospective Analysis
- APCRA Chemicals for Retrospective Analysis\_App\_List\_448\_Chemicals
- ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances
- ATSDR Toxic Substances Portal Chemical List
- Bisphenol Compounds
- California Office of Environmental Health Hazard Assessment
- Chemicals with interesting names
- CMAP
- DNT Screening Library
- Drinking Water Suspects, KWR Water, Netherlands
- EDSP Universe
- EPA Chemicals associated with hydraulic fracturing
- EPA Consumer Products Suspect Screening Results
- EPA Consumer Products Suspect Screening Results
- EPA Integrated Risk Information System (IRIS)
- EPA PFAS Cross-Agency Research List
- EPA PFAS List of 75 Test Samples (Set 1)
- EPACPJH - EPA Cell Painting Reference Chemical Set
- EPAHFR - EPA Chemicals associated with hydraulic fracturing
- EU Cosmetic Ingredients Inventory (Combined 2000/2006)

# Batch Search for SSA/NTA

## Batch Search



### Step One: Select Input

Please enter one identifier per line

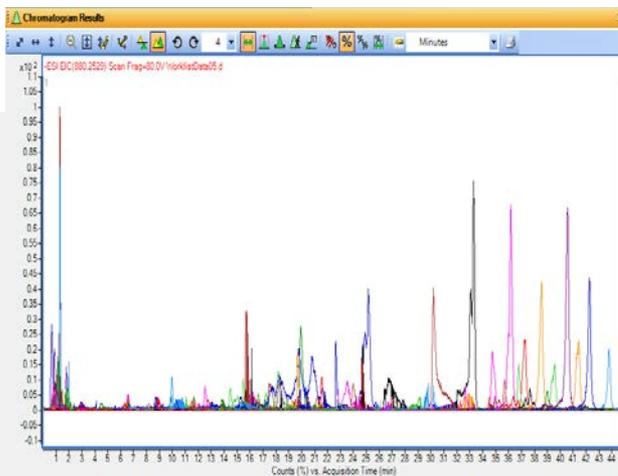
#### Select Input Type(s)

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

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

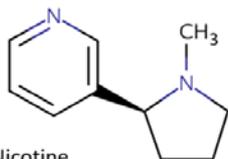
Chemical Data

C10H14N2

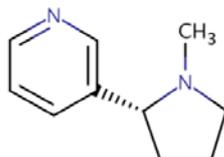


# MS-Ready Structures improve database searching

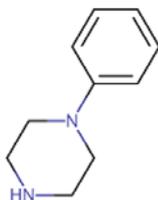
## Exact Formula Match and MS-Ready Match



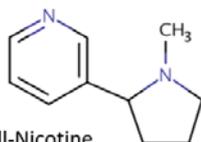
Nicotine  
DTXSID1020930 | DTXCID9028128  
Tox: yes | Expo: yes | Bioassay: yes  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub> | 54-11-5 | 87



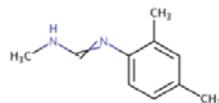
D-Nicotine  
DTXSID0046351 | DTXCID9028128  
Tox: no | Expo: yes | Bioassay: yes  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub> | 25162-00-9 | 21



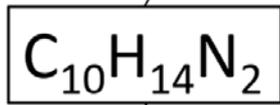
Phenylpiperazine  
DTXSID8057855 | DTXCID9031644  
Tox: no | Expo: no | Bioassay: yes  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub> | 25162-00-9 | 32



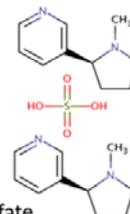
dl-Nicotine  
DTXSID3048154 | DTXCID9028128  
Tox: yes | Expo: no | Bioassay: yes  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub> | 22083-74-5 | 16



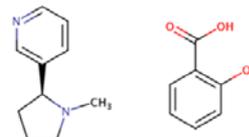
N'-(2,4-Dimethylphenyl)-N-methylformamide  
DTXSID1037696 | DTXCID9017696  
Tox: no | Expo: Yes | Bioassay: yes  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub> | 33089-74-6 | 27



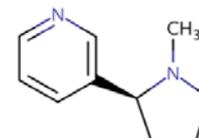
## MS-Ready Match Only



Nicotine sulfate  
DTXSID8021725 | DTXCID9028128  
Tox: yes | Expo: yes | Bioassay: yes  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub>.C<sub>10</sub>H<sub>14</sub>N<sub>2</sub>.SH<sub>2</sub>O<sub>4</sub> | 65-30-5 | 28



Benzoic acid, 2-hydroxy-, compd. with 3-((2S)-1-methyl-2-pyrrolidinyl)pyridine (1:1)  
DTXSID5075319 | DTXCID9028128 | DTXCID206368  
Tox: no | Expo: yes | Bioassay: no  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub>.C<sub>7</sub>H<sub>6</sub>O<sub>3</sub> | 29790-52-1 | 7



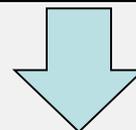
HCl  
Nicotine hydrochloride  
DTXSID6020931 | DTXCID9028128  
Tox: no | Expo: yes | Bioassay: yes  
C<sub>10</sub>H<sub>14</sub>N<sub>2</sub>.HCl | 2820-51-1 | 10

**LEGEND:** Preferred Name  
DTXSID | MS-ready DTXCID  
Avail. Data: Toxicity | Exposure | Bioassay  
Formula | CAS | Data Sources

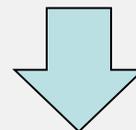
# Data Source Ranking for Identification in SSA/NTA

- Mass and/or formula unknown to a researcher, contained within a reference database
- Most likely candidate chemicals have the most references/sources

C<sub>14</sub>H<sub>22</sub>N<sub>2</sub>O<sub>3</sub>  
266.16304



Chemical  
Reference  
Database



Sorted  
candidate  
structures

RAPID COMMUNICATION

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

Andrew D. McEachran<sup>1</sup> · Jon R. Sobus<sup>2</sup> · Antony J. Williams<sup>3</sup>

- On same 162 chemicals, Dashboard outperforms ChemSpider

	Mass-based searching		Formula-based searching	
	Dashboard	ChemSpider	Dashboard	ChemSpider
Average rank position	1.3	2.2 <sup>a</sup>	1.2	1.4
Percent in #1 position	85%	70%	88%	80%

<sup>a</sup> Average rank in ChemSpider shown here does not include an outlier where the rank was 201, when added the average rank position is 3.5

# Increasing Data Streams to Improve Identifications

- US EPA CompTox Dashboard Data Sources (DS)
- PubChem Data Source Count
- PubMed Reference Count
- Presence in STOFF-IDENT Database
- Predicted Environmental Media Occurrence
- CPDat Product Occurrence Count
- OPERA PhysChem Properties
- NORMAN Network Priority List



PubChem

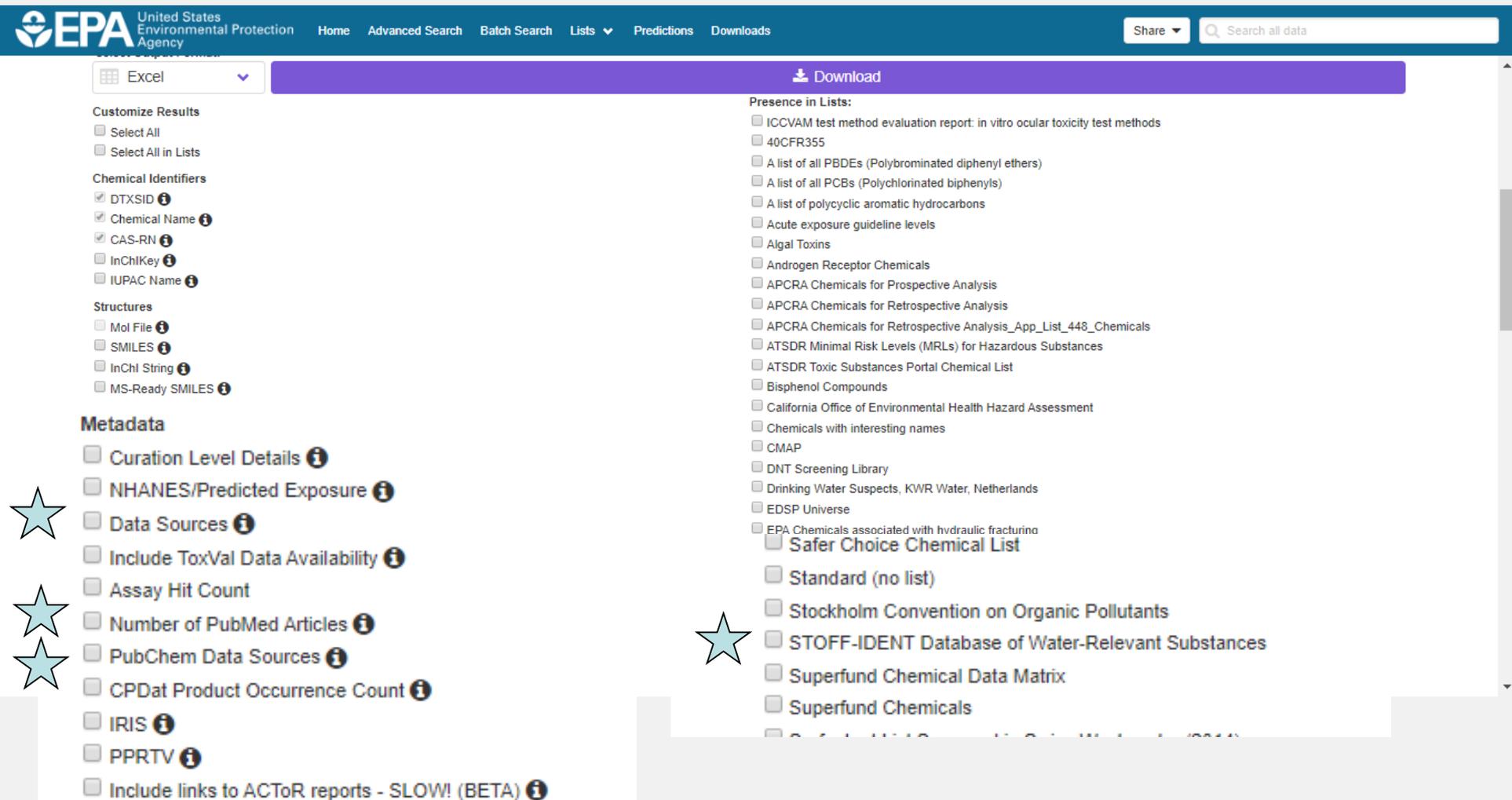


PubMed.gov



STOFF IDENT

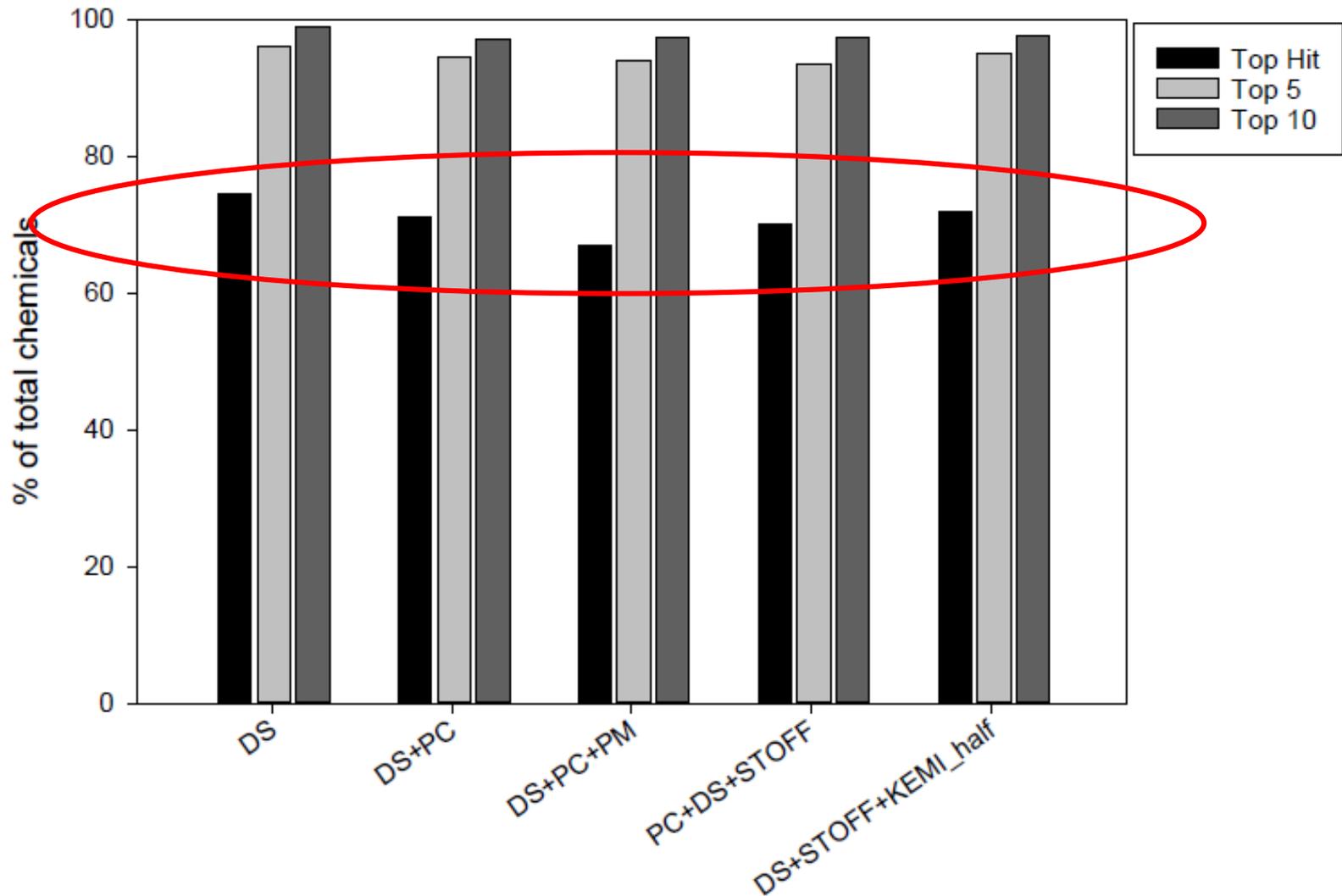
# All available via Batch Search:



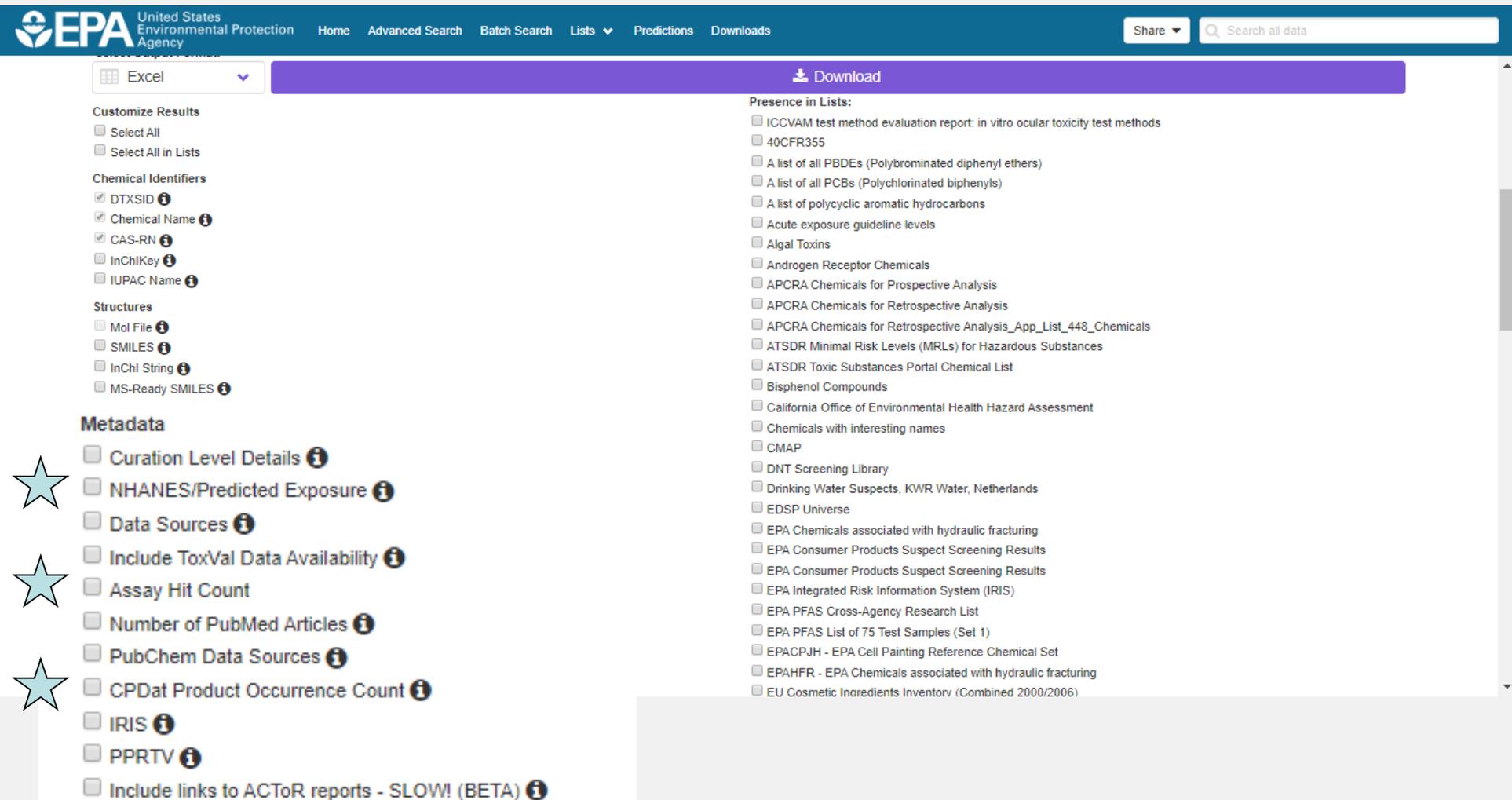
The screenshot shows the EPA Batch Search interface. At the top, there is a navigation bar with the EPA logo, the text "United States Environmental Protection Agency", and links for "Home", "Advanced Search", "Batch Search", "Lists", "Predictions", and "Downloads". A search bar on the right contains the text "Search all data". Below the navigation bar, there is a "Download" button and a "Share" dropdown menu. The main content area is divided into two columns. The left column contains several sections: "Customize Results" with checkboxes for "Select All" and "Select All in Lists"; "Chemical Identifiers" with checkboxes for "DTXSID", "Chemical Name", "CAS-RN", "InChIKey", and "IUPAC Name"; "Structures" with checkboxes for "Mol File", "SMILES", "InChI String", and "MS-Ready SMILES"; and "Metadata" with checkboxes for "Curation Level Details", "NHANES/Predicted Exposure", "Data Sources", "Include ToxVal Data Availability", "Assay Hit Count", "Number of PubMed Articles", "PubChem Data Sources", "CPDat Product Occurrence Count", "IRIS", "PPRTV", and "Include links to ACToR reports - SLOW! (BETA)". The right column contains a "Presence in Lists:" section with a long list of checkboxes for various lists, including "ICCVAM test method evaluation report: in vitro ocular toxicity test methods", "40CFR355", "A list of all PBDEs (Polybrominated diphenyl ethers)", "A list of all PCBs (Polychlorinated biphenyls)", "A list of polycyclic aromatic hydrocarbons", "Acute exposure guideline levels", "Algal Toxins", "Androgen Receptor Chemicals", "APCRA Chemicals for Prospective Analysis", "APCRA Chemicals for Retrospective Analysis", "APCRA Chemicals for Retrospective Analysis\_App\_List\_448\_Chemicals", "ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances", "ATSDR Toxic Substances Portal Chemical List", "Bisphenol Compounds", "California Office of Environmental Health Hazard Assessment", "Chemicals with interesting names", "CMAP", "DNT Screening Library", "Drinking Water Suspects, KWR Water, Netherlands", "EDSP Universe", "EPA Chemicals associated with hydraulic fracturing", "Safer Choice Chemical List", "Standard (no list)", "Stockholm Convention on Organic Pollutants", "STOFF-IDENT Database of Water-Relevant Substances", "Superfund Chemical Data Matrix", and "Superfund Chemicals". Three light blue stars are overlaid on the image: one next to "NHANES/Predicted Exposure" in the Metadata section, one next to "Safer Choice Chemical List" in the Presence in Lists section, and one next to "Stockholm Convention on Organic Pollutants" in the Presence in Lists section.

# Identification ranks for 1783 chemicals using multiple data streams

$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \dots$$



# Linking exposure potential and bioactivity for prioritization



The screenshot shows the EPA website's search results page. At the top, there is a navigation bar with the EPA logo, the text "United States Environmental Protection Agency", and links for "Home", "Advanced Search", "Batch Search", "Lists", "Predictions", and "Downloads". On the right side of the navigation bar, there is a "Share" button and a search input field containing "Search all data".

Below the navigation bar, there is a purple bar with a "Download" button. To the left of this bar, there is a dropdown menu showing "Excel".

The main content area is divided into several sections:

- Customize Results:** Includes checkboxes for "Select All", "Select All in Lists", "DTXSID", "Chemical Name", "CAS-RN", "InChIKey", and "IUPAC Name".
- Chemical Identifiers:** Includes checkboxes for "Mol File", "SMILES", "InChI String", and "MS-Ready SMILES".
- Structures:** Includes checkboxes for "Mol File", "SMILES", "InChI String", and "MS-Ready SMILES".
- Metadata:** Includes checkboxes for "Curation Level Details", "NHANES/Predicted Exposure", "Data Sources", "Include ToxVal Data Availability", "Assay Hit Count", "Number of PubMed Articles", "PubChem Data Sources", "CPDat Product Occurrence Count", "IRIS", "PPRTV", and "Include links to ACToR reports - SLOW! (BETA)".
- Presence in Lists:** Includes checkboxes for "ICCVAM test method evaluation report: in vitro ocular toxicity test methods", "40CFR355", "A list of all PBDEs (Polybrominated diphenyl ethers)", "A list of all PCBs (Polychlorinated biphenyls)", "A list of polycyclic aromatic hydrocarbons", "Acute exposure guideline levels", "Algal Toxins", "Androgen Receptor Chemicals", "APCRA Chemicals for Prospective Analysis", "APCRA Chemicals for Retrospective Analysis", "APCRA Chemicals for Retrospective Analysis\_App\_List\_448\_Chemicals", "ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances", "ATSDR Toxic Substances Portal Chemical List", "Bisphenol Compounds", "California Office of Environmental Health Hazard Assessment", "Chemicals with interesting names", "CMAP", "DNT Screening Library", "Drinking Water Suspects, KWR Water, Netherlands", "EDSP Universe", "EPA Chemicals associated with hydraulic fracturing", "EPA Consumer Products Suspect Screening Results", "EPA Consumer Products Suspect Screening Results", "EPA Integrated Risk Information System (IRIS)", "EPA PFAS Cross-Agency Research List", "EPA PFAS List of 75 Test Samples (Set 1)", "EPACPJH - EPA Cell Painting Reference Chemical Set", "EPAHFR - EPA Chemicals associated with hydraulic fracturing", and "EU Cosmetic Ingredients Inventory (Combined 2000/2006)".

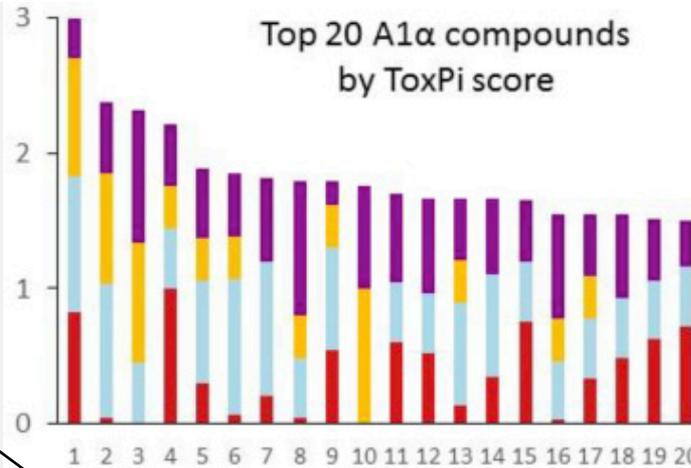
# Suspect screening and non-targeted analysis of drinking water using point-of-use filters ☆

Seth R. Newton <sup>a</sup>, Rebecca L. McMahan <sup>a, b</sup>, Jon R. Sobus <sup>a</sup>, Kamel Mansouri <sup>b, c, 1</sup>, Antony J. Williams <sup>c</sup>, Andrew D. McEachran <sup>b, c</sup>, Mark J. Strynar <sup>a</sup>

### Metadata

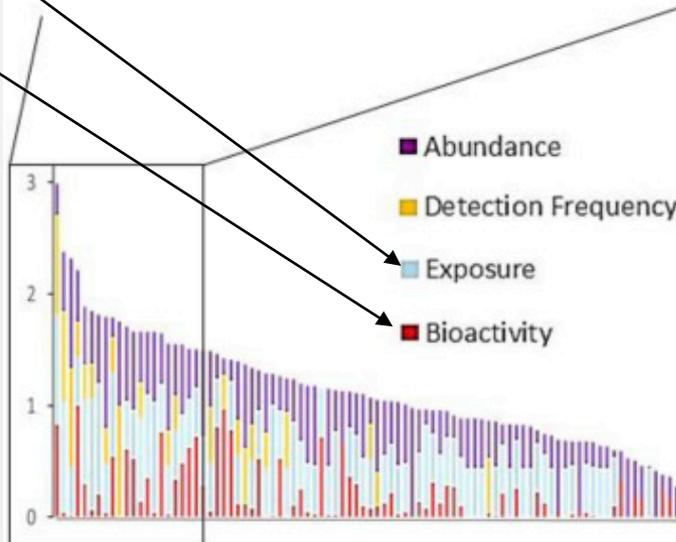
- Curation Level Details ⓘ
- NHANES/Predicted Exposure ⓘ
- Data Sources ⓘ
- Include ToxVal Data Availability ⓘ
- Assay Hit Count
- Number of PubMed Articles ⓘ
- PubChem Data Sources ⓘ
- CPDat Product Occurrence Count ⓘ
- IRIS ⓘ
- PPRTV ⓘ
- Include links to ACToR reports - SLOW! (BETA) ⓘ

Top 20 A1α compounds by ToxPi score



#	Compound	ToxPi Score
1	1,2-Benzisothiazolin-3-one*	2.99
2	Diethylene glycol	2.38
3	N-[3-(Dimethylamino)propyl] methacrylamide	2.32
4	Nonylparaben	2.22
5	Dipentyl phthalate	1.89
6	2-[2-(2-Butoxyethoxy) ethoxy]ethanol*	1.85
7	N,N-Dimethyldodecan-1-amine*	1.81
8	Sucralose	1.80
9	PFOS*	1.79
10	2-(2-Ethoxyethoxy) ethyl acetate*	1.76
11	TDCPP*	1.71
12	Zearalanol	1.67
13	PFOA*	1.66
14	Butylparaben	1.66
15	Noristerat	1.65
16	p-Syneprine	1.55
17	Alprostadiol	1.55
18	Sciarelol	1.55
19	PFDA*	1.51
20	Simvastatin	1.50

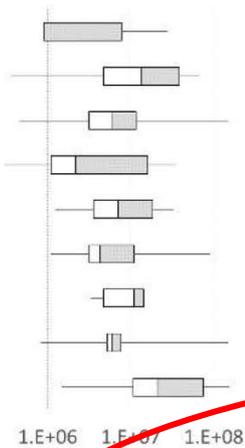
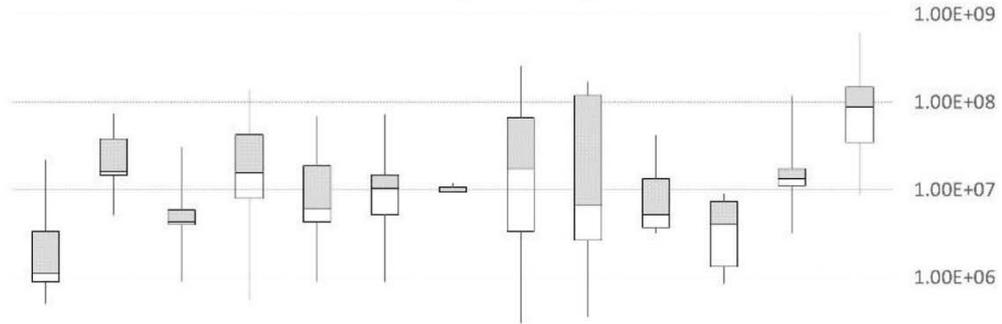
\*Confirmed with standard



All A1α compounds

# CPCat terms can inform source

### B) Chromatographic Peak Area Range In Categories



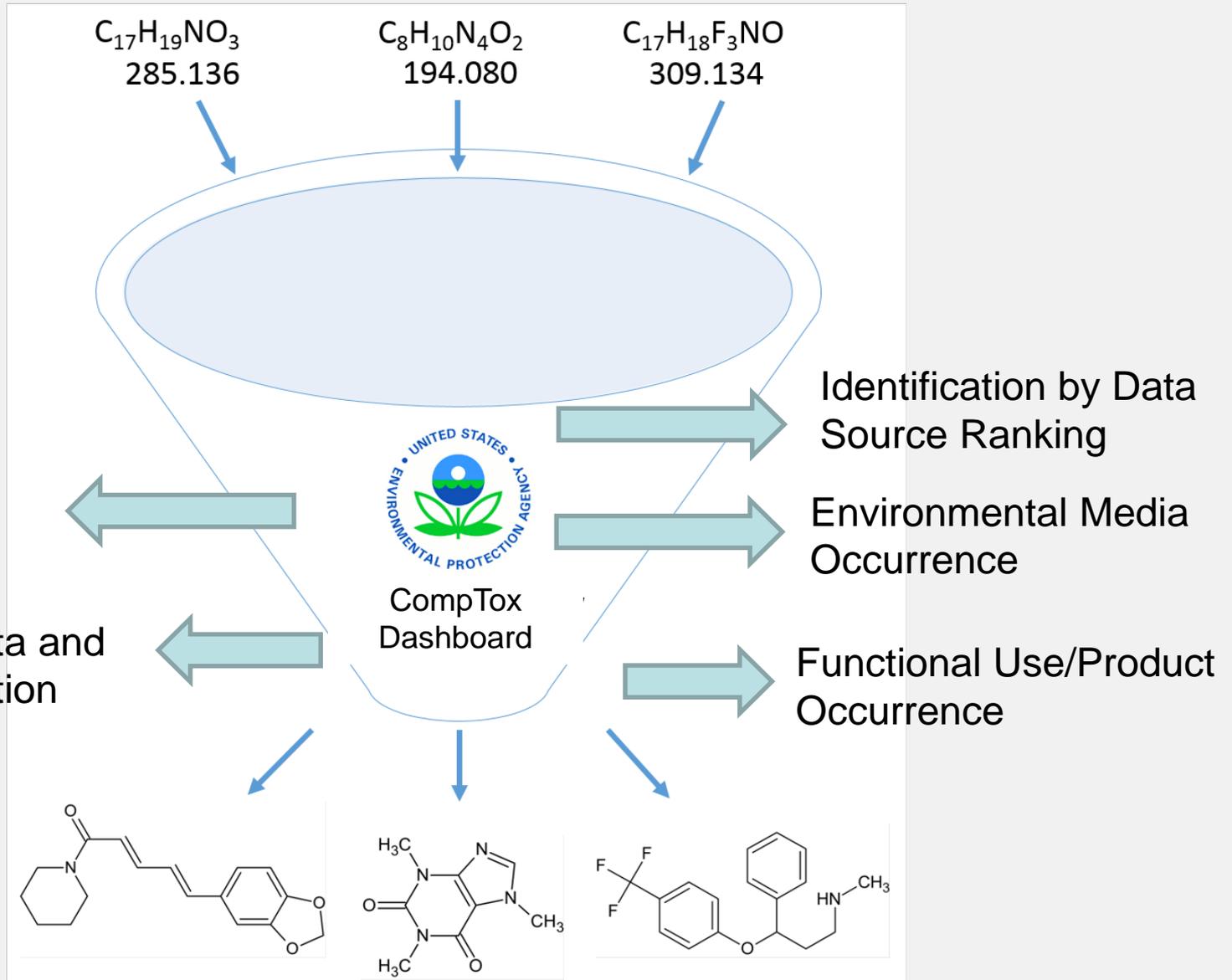
	9	10	13	16	22	28	9	26	8	9	3	18	66	Total
8	1	1	1	2	1	1		1						Durham (Tap 1)
34	2	4	2	4	4	4		4	1	2		4	3	Durham (Tap 2)
47	1		2	1	4	4		4	2	1	1	3	24	Chapel Hill (Well)
17	1	2	1	2	1	3		1	1			1	4	Chapel Hill (Tap)
23	1		2		4	5		1	1	1	1	3	4	Apex (Tap)
18	1		2		2	3		4		1		1	4	Cary (Tap)
13	1	3		3				1				2	3	Raleigh (Tap)
28	1		2	1	2	3	4			1		1	13	Pittsboro (Well)
49			1	3	4	5	5	10	3	3	1	3	11	Pittsboro (Tap)
Total														
Antimicrobial														
Herbicide														
Pesticide Active And Consumer														
Pesticide Active No Consumer														
Pesticide Inert														
Consumer and Industrial Process														
Consumer No Industrial Process														
Industrial Process No Consumer														
Food Additive														
Colorant														
Fragrance														
Personal Care Product														
Other														

### C) Chromatographic Peak Area Range In Samples

### D) Number of Chemicals in Each Category by Sample



# Dashboard in SSA/NTA Workflows

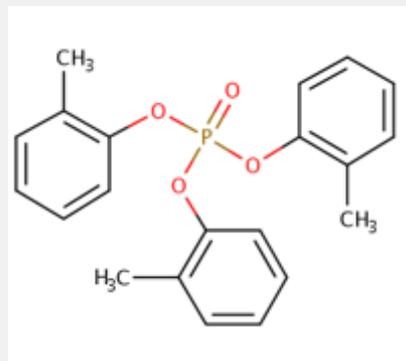
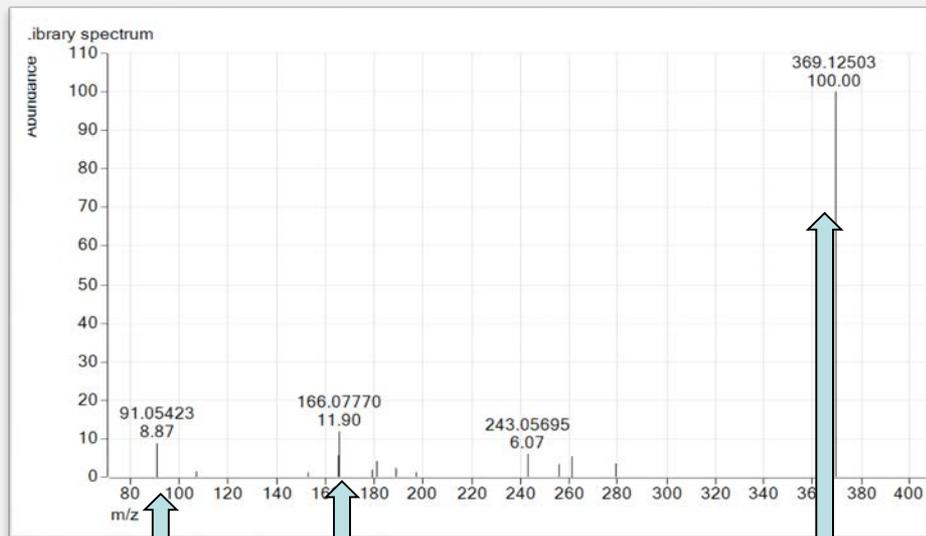


# Future Directions

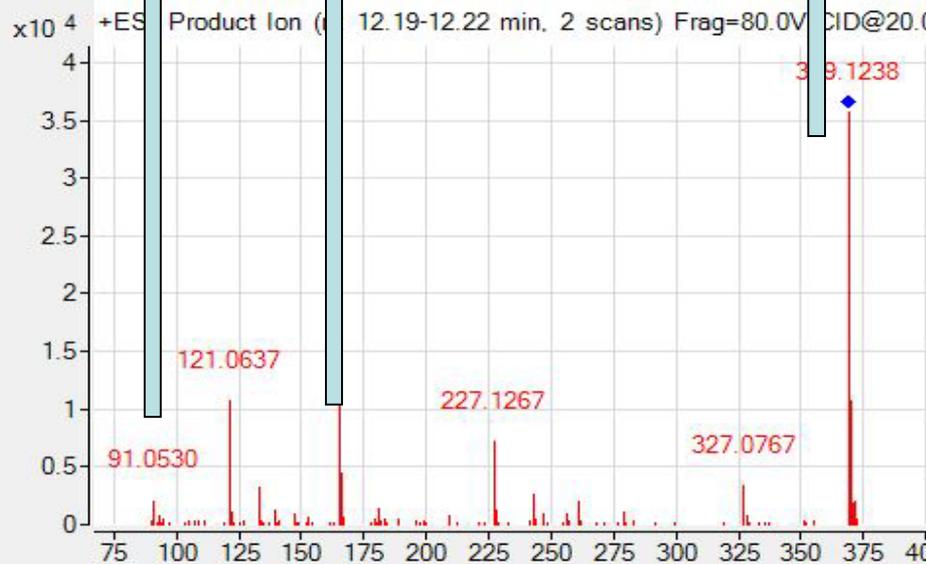
- Media occurrence prediction models
- Combined data visualization
- MS/MS match score calculations via the Dashboard
- Retention time index (RTI) predictions
- Ongoing expansion of the database
- Integration to public MS databases

# MS/MS Spectral Matching for Identification

**Library  
Fragmentation  
Spectra (20eV)**

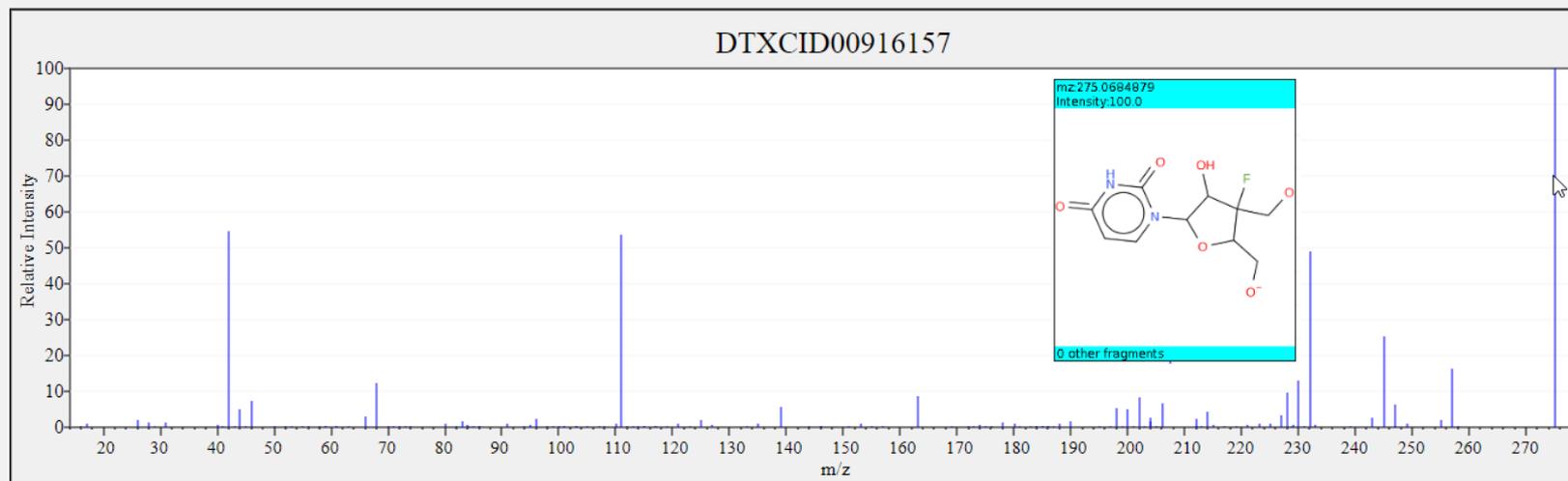


**Observed  
Fragmentation  
Spectra (20eV)**



➔ **Match  
Score**

# Predicted MS/MS Spectral Matching for Identification



## CASMI Contest Challenge Set (n=208)

### CFM-ID only

	# Identified	% of Total
#1 Hits	89	43%
Top 5	154	74%
Top 10	174	84%
Top 20	190	91%

### CFM-ID +DSSTox Data Sources

	# Identified	% of Total
#1 Hits	154	74%
Top 5	195	94%
Top 10	198	95%
Top 20	202	97%

# Conclusions

- CompTox Dashboard provides access to a wealth of chemistry data that can be leveraged for exposure analyses
- NTA/SSA using the Dashboard results in high performance of identifications
- New data streams and predictions linked within the Dashboard further enhance NTA/SSA capabilities

# Acknowledgements

## EPA NCCT

Tony Williams  
Chris Grulke  
Jeff Edwards

## EPA NERL

Katherine Phillips  
Kristin Isaacs  
Kathie Dionisio  
Jon Sobus  
Mark Strynar  
Elin Ulrich  
Seth Newton  
Hussein Al-Ghoul\*

## External Collaborators

Emma Schymanski- Univ.  
Luxembourg  
Christoph Ruttkies- IPB,  
Halle  
Kamel Mansouri- ILS, Inc

\*ORAU Participant

# Questions?

- [mceachran.andrew@epa.gov](mailto:mceachran.andrew@epa.gov)
- <http://orcid.org/0000-0003-1423-330X>
- Associated presentations:
  - ANYL 100: Developing tools for high resolution mass spectrometry-based screening via the EPA's CompTox Chemistry Dashboard
  - ENVR 152: EPA Comptox Chemistry Dashboard as a data integration hub for environmental chemistry data
  - AGRO 107: Consensus ranking and fragmentation prediction for identification of unknowns in high resolution mass spectrometry