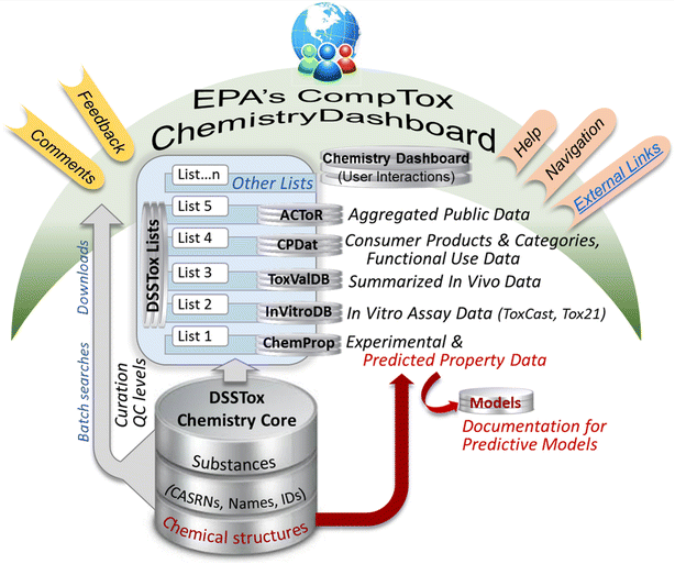
http://www.orcid.org/0000-0002-2668-4821

EPA United States Environmental Protection Agency

The EPA Comptox Chemistry Dashboard: A Web-Based Data Integration Hub for Environmental Chemistry and Toxicology Data



Antony Williams

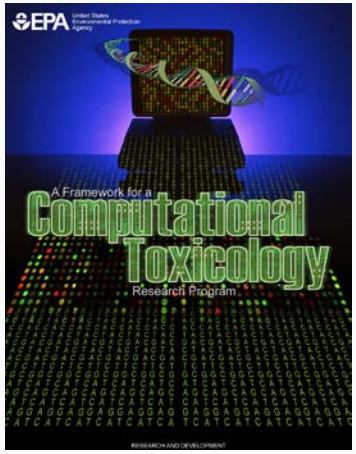
U.S. Environmental Protection Agency, RTP, NC

This work was reviewed by the U.S. EPA and approved for presentation but does not necessarily reflect official Agency policy.

April 17th 2018

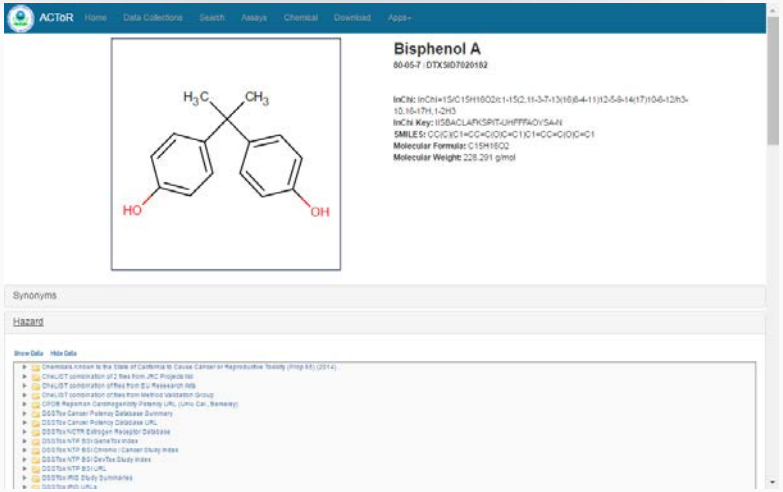
SETAC North America Focused Topic Meeting

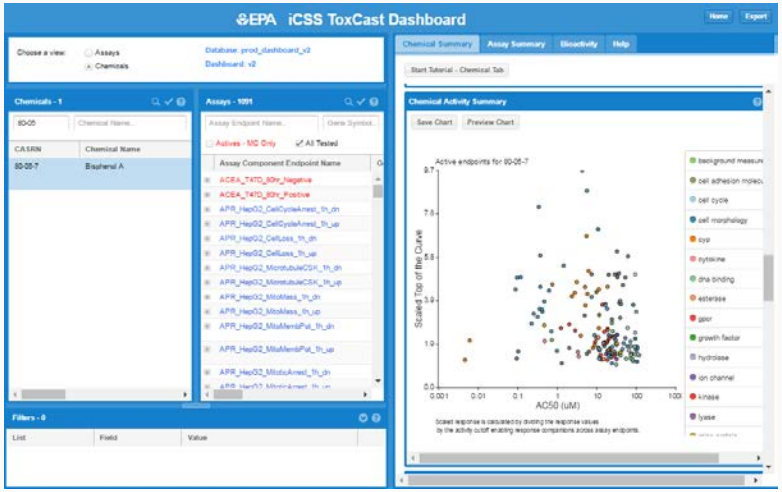
National Center for Computational Toxicology

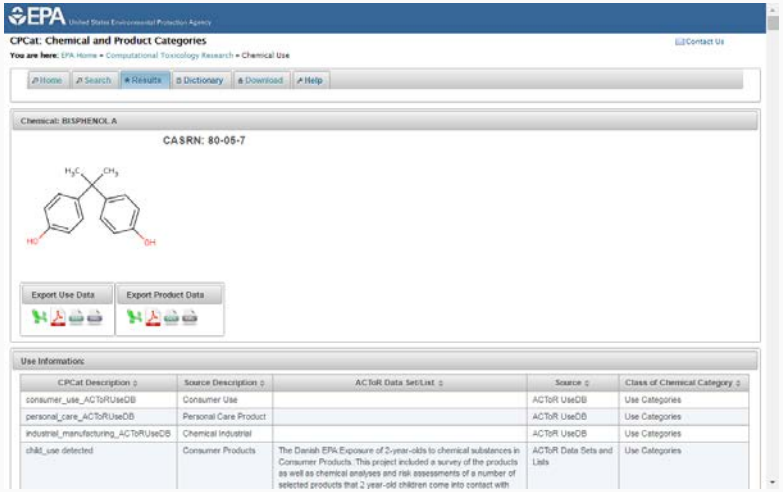


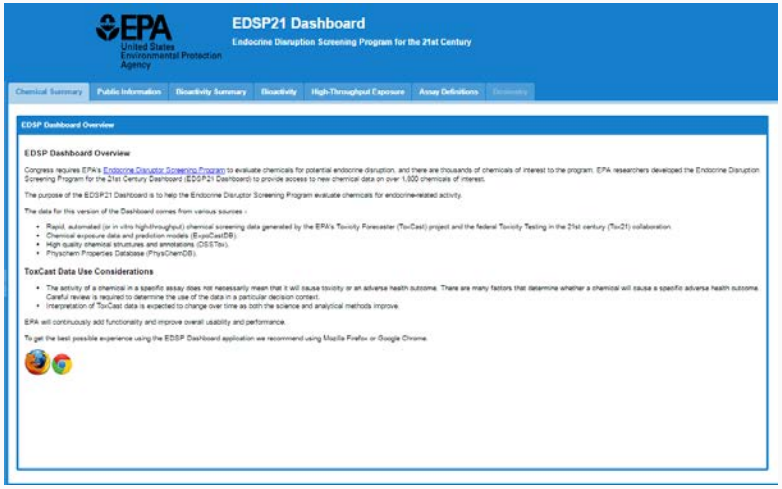
- National Center for Computational Toxicology established in 2005 to integrate:
  - High-throughput and high-content technologies
  - Modern molecular biology
  - Data mining and statistical modeling
  - Computational biology and chemistry
- NCCT outputs: include a lot of data, models, algorithms and software applications
- We produce Open Data – we want people to interrogate it, learn from it, develop understanding

Early Dashboard Applications



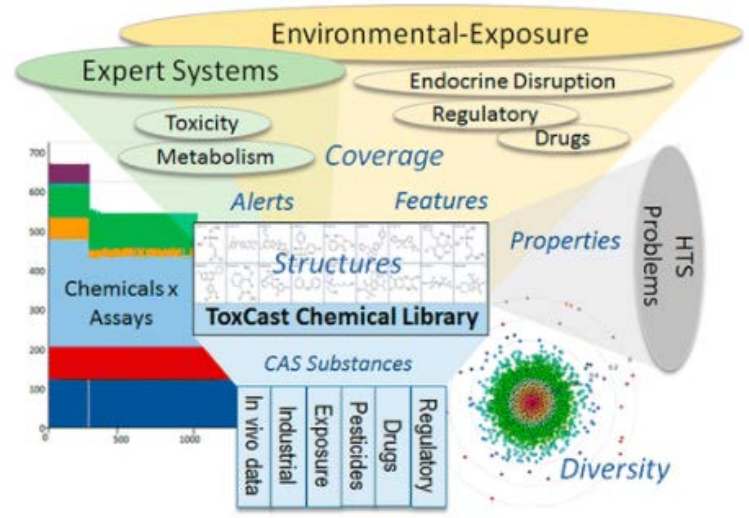






Integrating All of Our Data

- Breadth of data has grown over the years
  - *In vitro* high-throughout screening data – thousands of chemicals across hundreds of assays
  - Hazard data – human and ecological
  - Exposure Data (exp. & pred.)
  - Predictive models
  - Aggregated data
    - Other EPA databases
    - International databases
    - Literature data



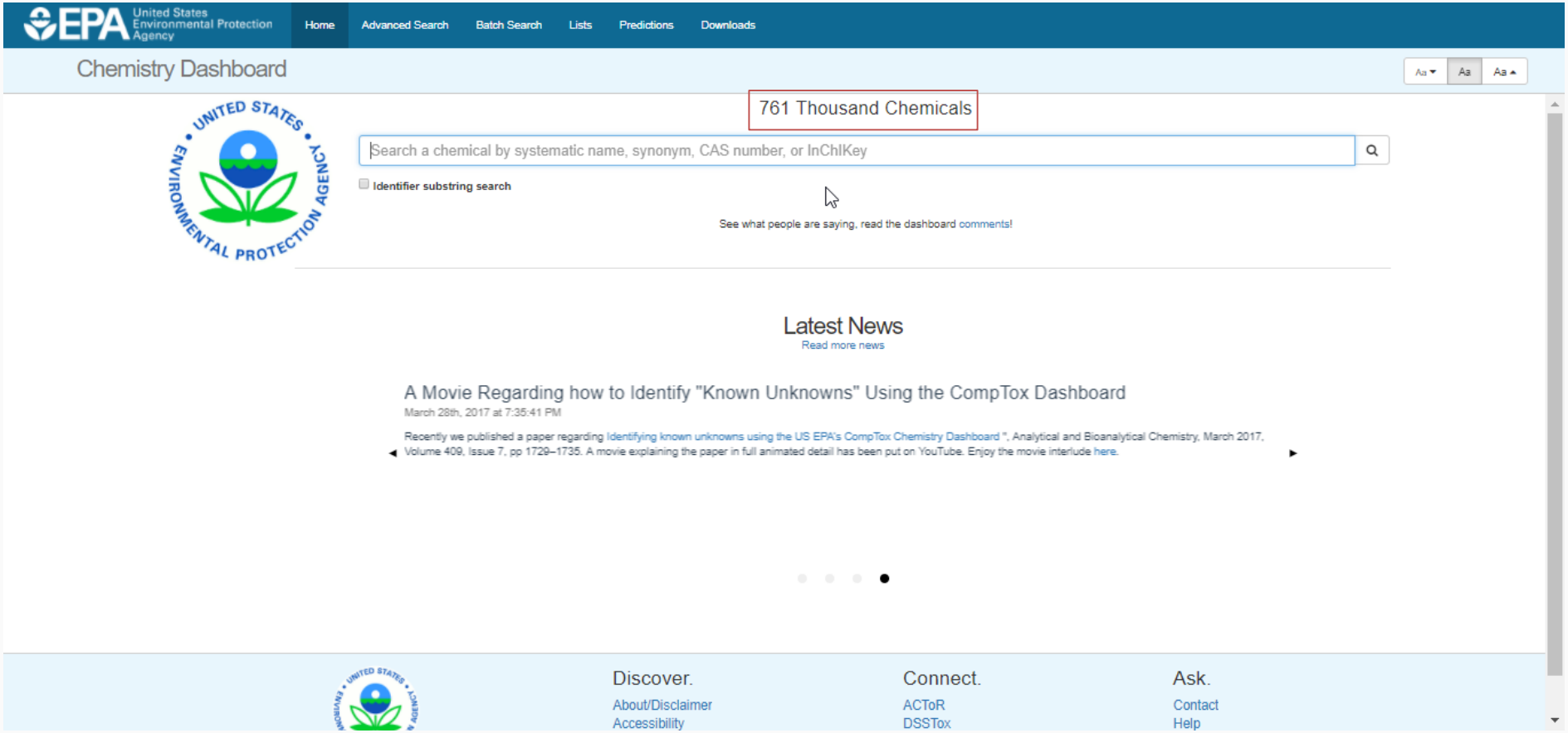
The CompTox Chemistry Dashboard

- A new architecture bringing together all data

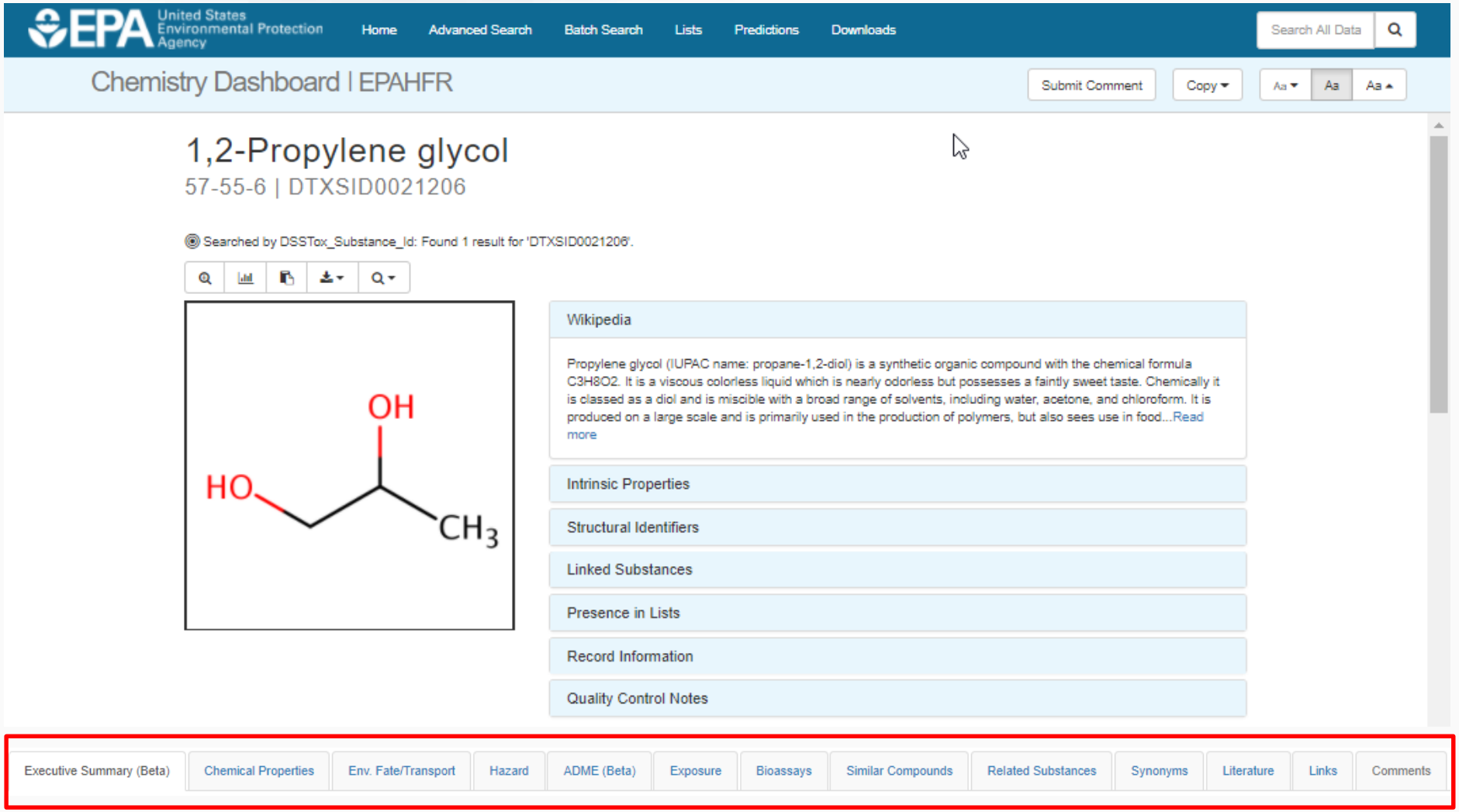
- A **publicly accessible website** delivering access:
  - ~760,000 chemicals with related property data
  - Experimental and predicted physicochemical property data
  - Experimental Human and Ecological hazard data
  - Integration to “biological assay data” for 1000s of chemicals
  - Information regarding consumer products containing chemicals
  - Links to other agency websites and public data resources
  - “Literature” searches for chemicals using public resources
  - “Batch searching” for thousands of chemicals
  - Real time prediction of physchem and toxicity endpoints

CompTox Chemistry Dashboard

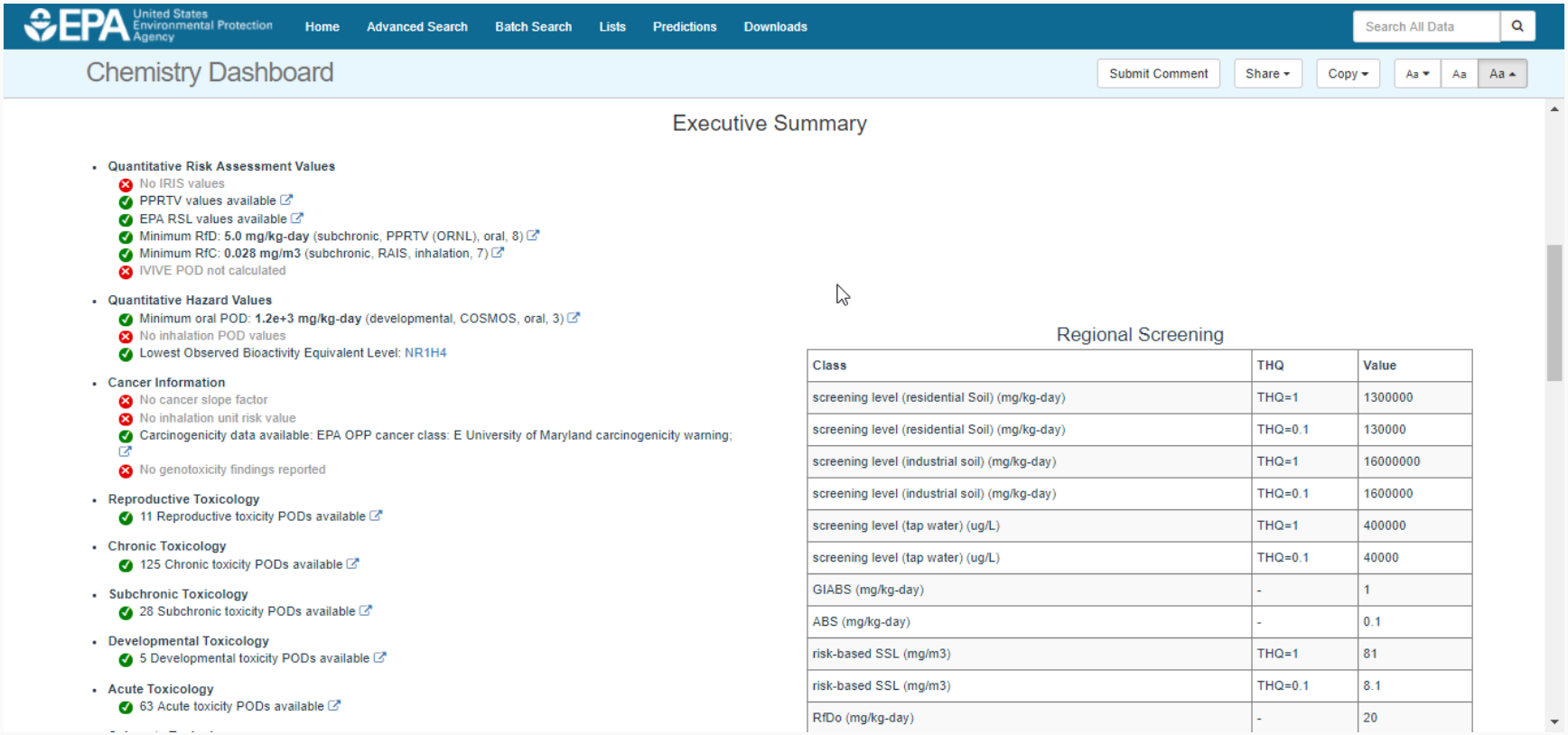
https://comptox.epa.gov/dashboard



Detailed Chemical Pages



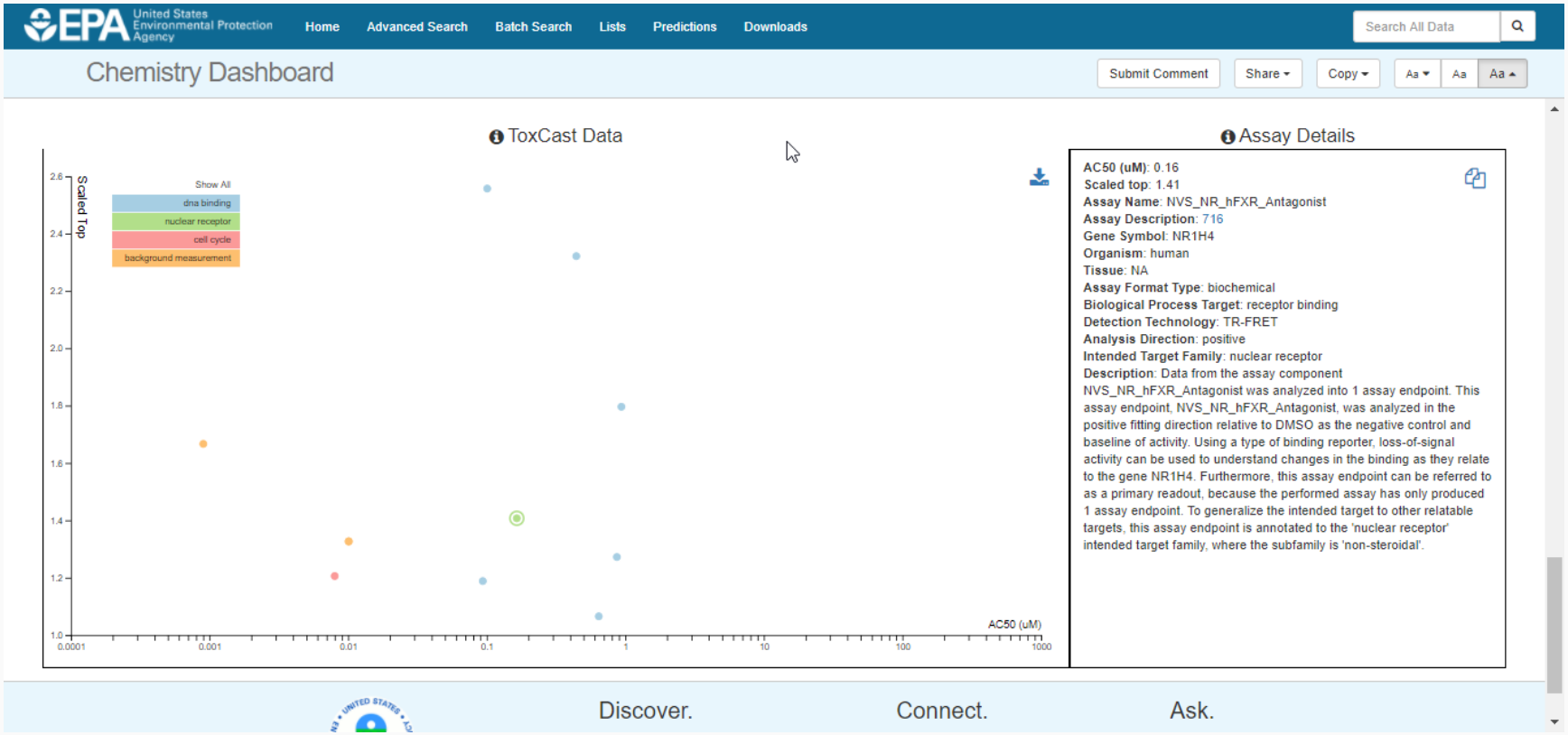
The Executive Summary



The Executive Summary

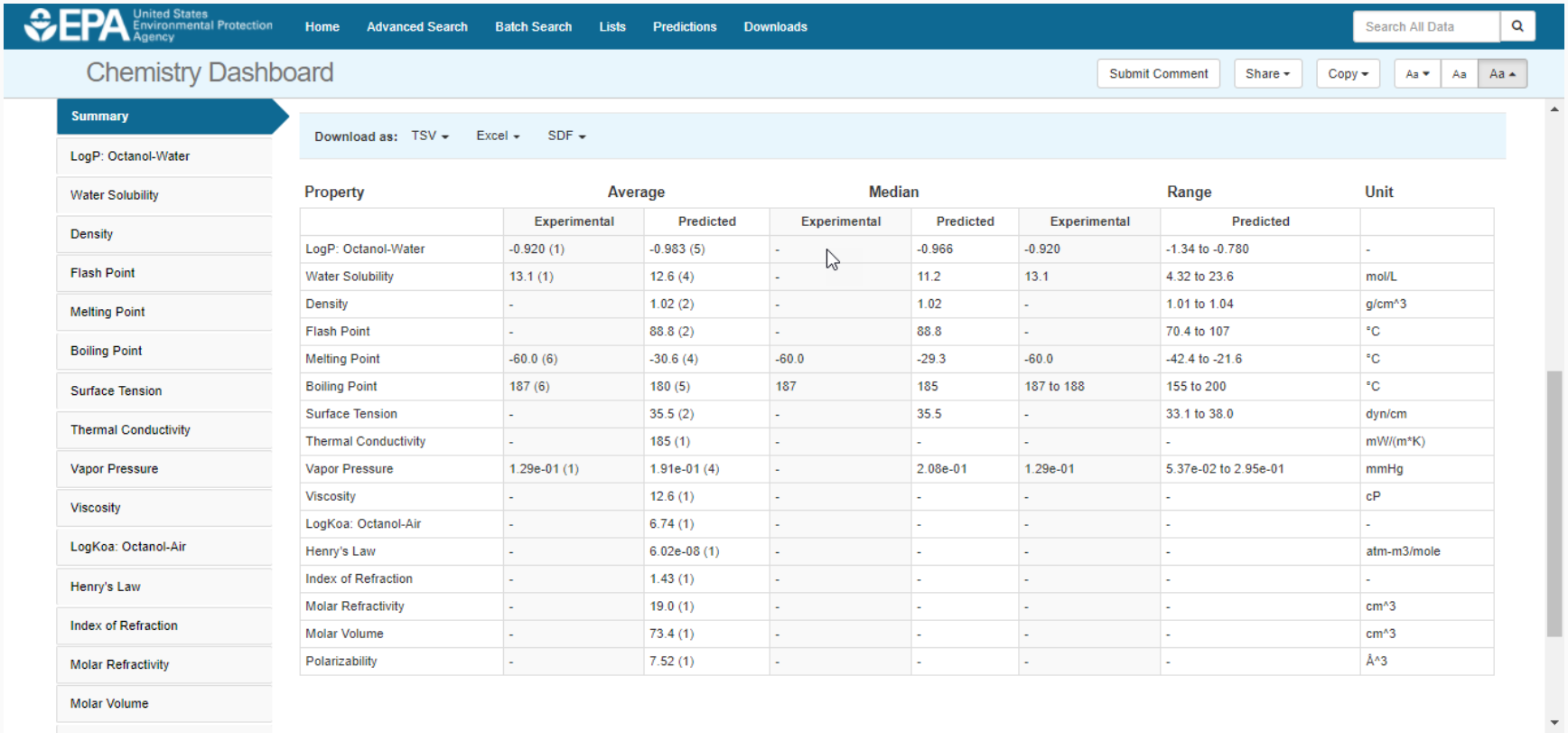
Executive Summary (Beta) | Chemical Properties | Env. Fate/Transport | Hazard | ADME (Beta) | Exposure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

The Executive Summary



Executive Summary (Beta) | Chemical Properties | Env. Fate/Transport | Hazard | ADME (Beta) | Exposure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

Properties, Fate and Transport



Property	Average		Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	-0.920 (1)	-0.983 (5)	-	-0.966	-0.920	-1.34 to -0.780	-
Water Solubility	13.1 (1)	12.6 (4)	-	11.2	13.1	4.32 to 23.6	mol/L
Density	-	1.02 (2)	-	1.02	-	1.01 to 1.04	g/cm^3
Flash Point	-	88.8 (2)	-	88.8	-	70.4 to 107	°C
Melting Point	-60.0 (6)	-30.6 (4)	-60.0	-29.3	-60.0	-42.4 to -21.6	°C
Boiling Point	187 (6)	180 (5)	187	185	187 to 188	155 to 200	°C
Surface Tension	-	35.5 (2)	-	35.5	-	33.1 to 38.0	dyn/cm
Thermal Conductivity	-	185 (1)	-	-	-	-	mW/(m*K)
Vapor Pressure	1.29e-01 (1)	1.91e-01 (4)	-	2.08e-01	1.29e-01	5.37e-02 to 2.95e-01	mmHg
Viscosity	-	12.6 (1)	-	-	-	-	cP
LogKoa: Octanol-Air	-	6.74 (1)	-	-	-	-	-
Henry's Law	-	6.02e-08 (1)	-	-	-	-	atm-m3/mole
Index of Refraction	-	1.43 (1)	-	-	-	-	-
Molar Refractivity	-	19.0 (1)	-	-	-	-	cm^3
Molar Volume	-	73.4 (1)	-	-	-	-	cm^3
Polarizability	-	7.52 (1)	-	-	-	-	Å^3

OPERA: OPEN Data and OPEN Models

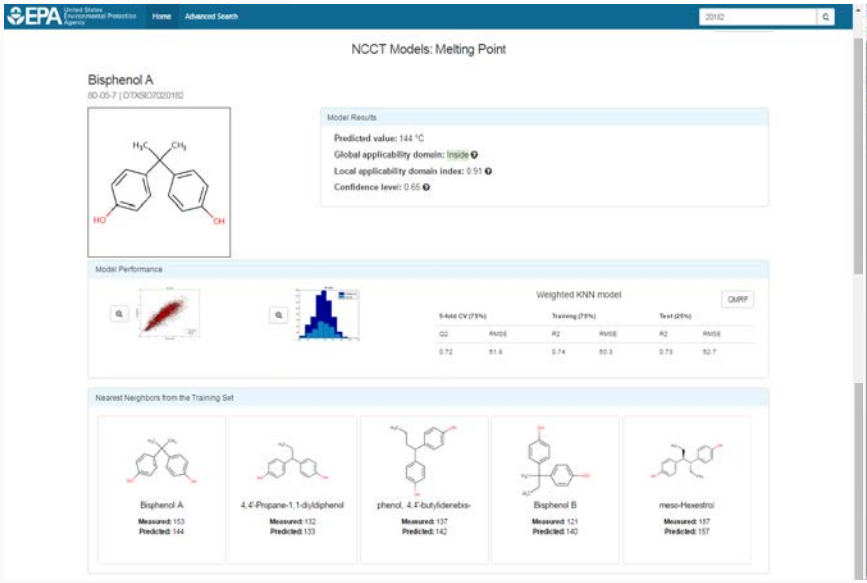
Mansouri *et al. J Cheminform (2018) 10:10*
https://doi.org/10.1186/s13321-018-0263-1

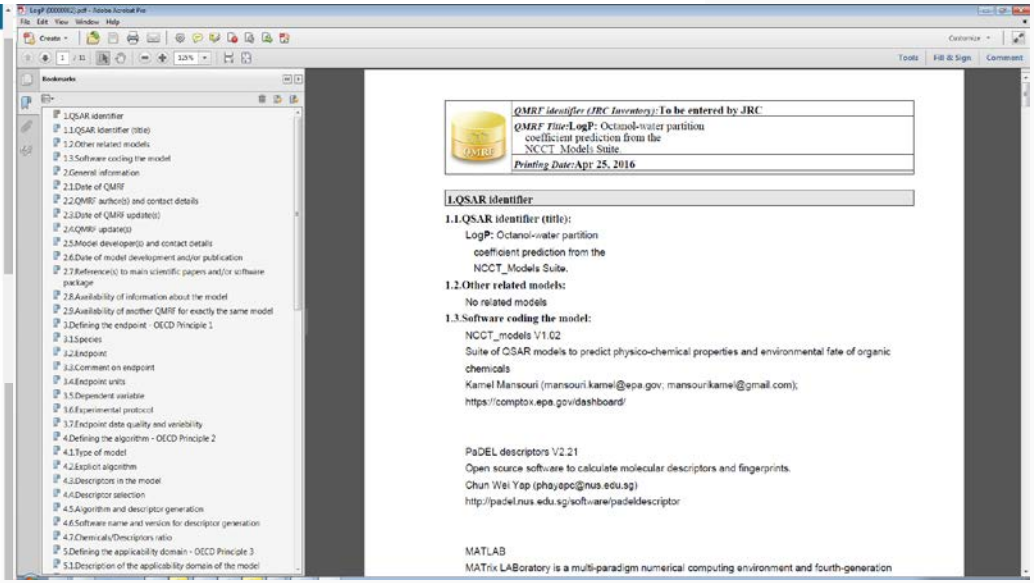
Journal of Cheminformatics

RESEARCH ARTICLE Open Access

OPERA models for predicting physicochemical properties and environmental fate endpoints

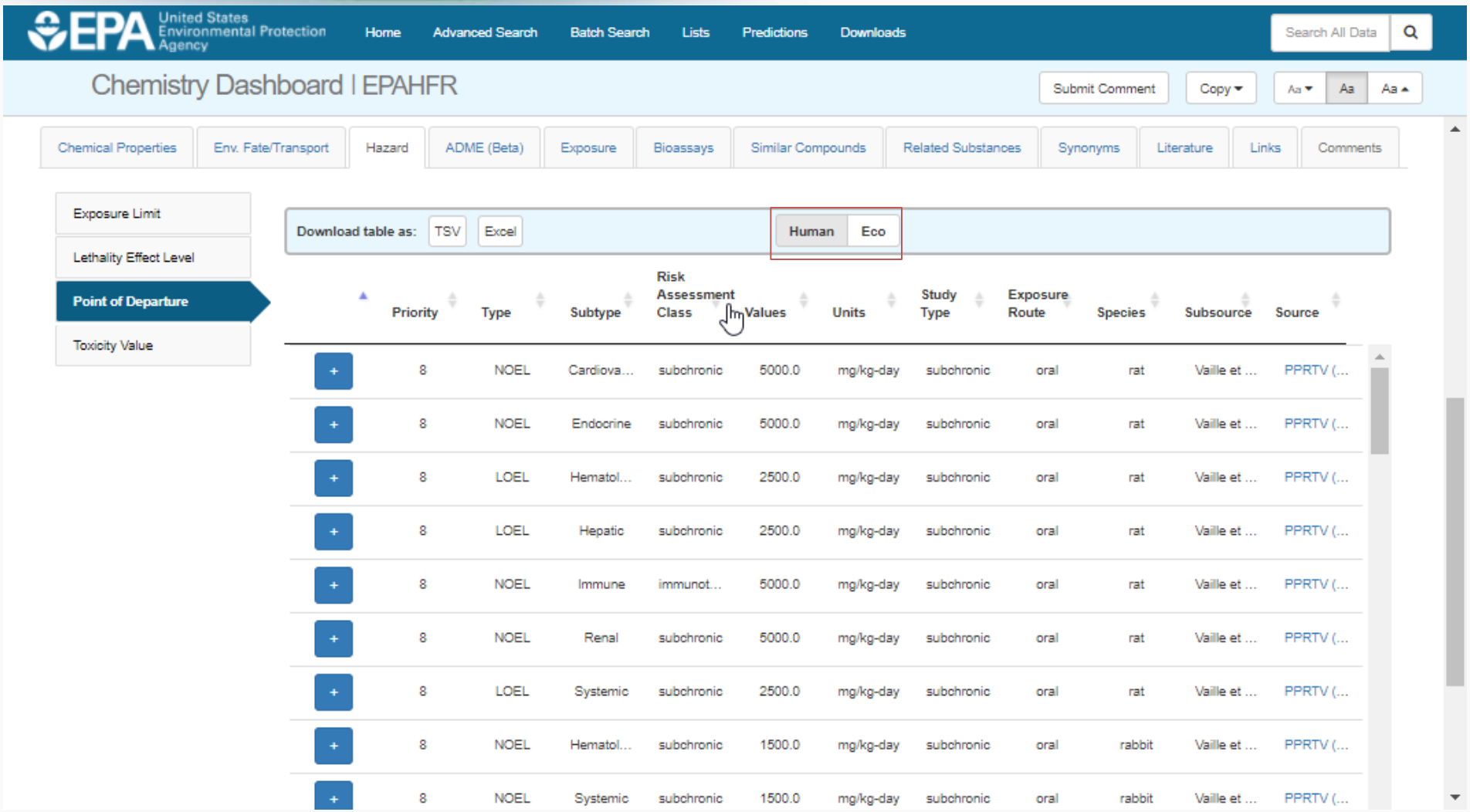
Kamel Mansouri[1,2,3*], Chris M. Grulke[1], Richard S. Judson[1] and Antony J. Williams[1]





Executive Summary (Beta) | Chemical Properties | Env. Fate/Transport | Hazard | ADME (Beta) | Exposure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

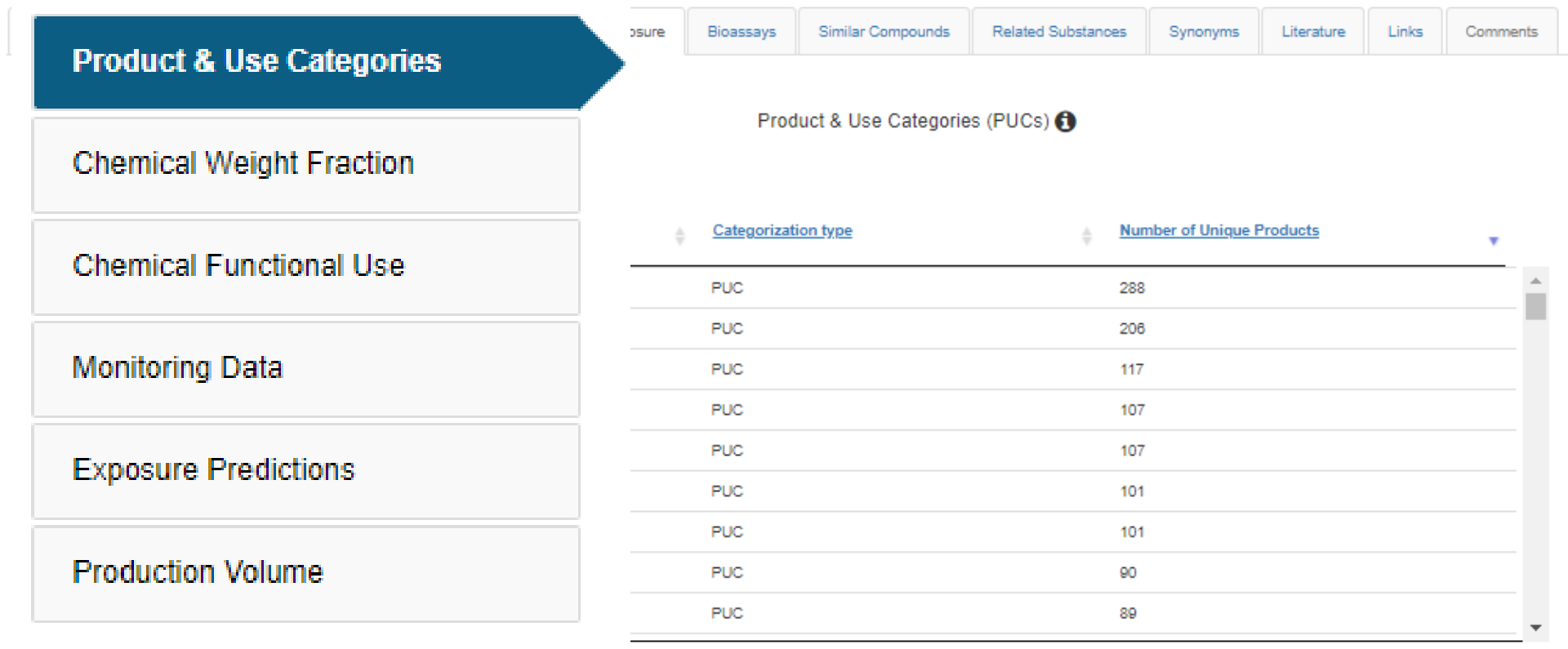
Access to Chemical Hazard Data



	Priority	Type	Subtype	Risk Assessment Class	Values	Units	Study Type	Exposure Route	Species	Subsource	Source
+	8	NOEL	Cardiova...	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Endocrine	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Hematol...	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Hepatic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Immune	immunot...	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Renal	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Systemic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Hematol...	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaille et ...	PPRTV (...
+	8	NOEL	Systemic	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaille et ...	PPRTV (...

Sources of Exposure to Chemicals

EPA United States Environmental Protection Agency



Product Composition Details

EPA United States Environmental Protection Agency

- Product & Use Categor...
- Chemical Weight Fra...
- Chemical Functional Use
- Monitoring Data
- Exposure Predictions

Chemical Weight Fractions

Download as: TSV | Excel

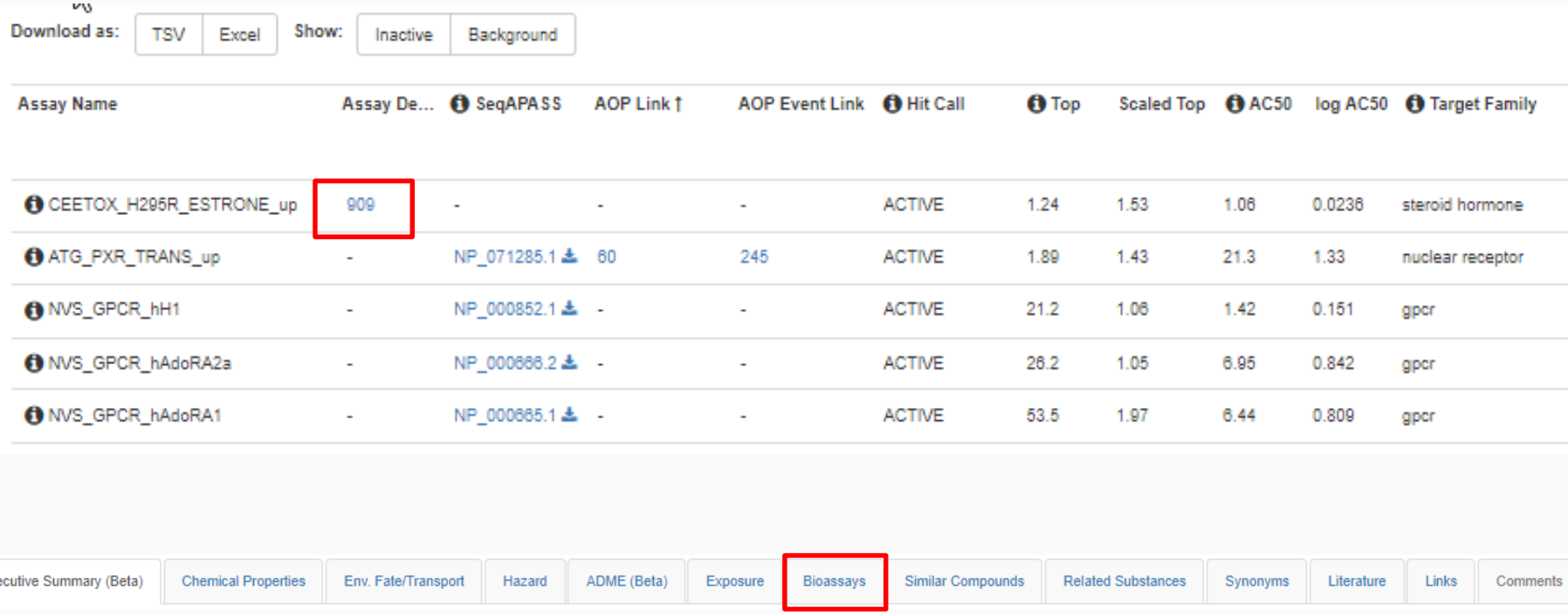
Product Name	Product Use Category	Minimum Weight Fraction	Maximum Weight Fraction	Data Type	Source
10-02199- calico tip & ...	personal care: nail poli...	0.01	0.05	MSDS	Retail Product Categori...
6095-6096 minwax wo...	home maintenance: fin...	0.01	0.01	MSDS	Retail Product Categori...
6095/6096 minwax wo...	home maintenance: fin...	0.01	0.01	MSDS	Retail Product Categori...
ab artificial nail remove...	personal care: nail poli...	0.01	0.03	MSDS	Retail Product Categori...
artificial nail remover 7...	personal care: nail poli...	0.01	0.05	MSDS	Retail Product Categori...
calico tip & glue remov...	personal care: nail poli...	0.01	0.05	MSDS	Retail Product Categori...
citristrip canadian strip...	home maintenance: str...	0.65	0.7	MSDS	Retail Product Categori...
citristrip stripping gel q...	home maintenance: str...	0.4	0.55	MSDS	Retail Product Categori...
citristrip stripping gel q...	home maintenance: str...	-	-	MSDS	Retail Product Categori...

In Vitro Bioassay Screening

ToxCast and Tox21

The Assay Table

- Detailed Assay Descriptions (growing list)
- Details in later talk…



Detailed Assay Description as PDF (limited number of assays)

NVS_NR_hFXR_Antagonist

Assay Title: NovaScreen Human Farnesoid x Receptor Alpha (FXR) Ligand-Binding Antagonist Screening Assay

1. Assay Descriptions

1.1. Overview

Assay Summary:

High-throughput screening of in vitro chemical-target interactions across a wide variety of compounds through a broad range of biochemical interactions will help describe the chemical-assay bioactivity space for chemicals with limited available information. There exists a large number of environmental chemicals for which there is little information about the potential for bioactivity. The NVS NR human farnesoid x receptor (FXR, NR1H4) agonist assay format allows for an efficient screening of thousands of chemicals for the ability to competitively bind to the ligand-binding domain of a xenobiotic sensing nuclear receptor. This assay was developed to screen the ToxCast chemical library for potential farnesoid x receptor ligand-binding activity using a TR-FRET competitive displacement assay and a known FXR receptor agonist (Chenodeoxycholic Acid, CDCA) as a reference compound. Biochemical high-throughput screening offers preliminary evidence for chemical targets in a cell or tissues which, when combined with information from literature or targeted in vivo studies, can indicate potential pathways for toxicity. This assay was run for a test duration of 1 hour in a 384-well plate.

1.2. Assay Definition

Assay Throughput:

Human FXR ligand-binding domain (LBD) incubated in 384-well microtiter plates for 1 hour prior to measuring ligand dependent binding of cofactor to the receptor using TR-FRET.

The Assay Table

- SeqaPASS files - Sequence Alignment to Predict Across Species Susceptibility

Download as: TSV Excel Show: Inactive Background

Assay Name	Assay De...	SeqAPASS	AOP Link †	AOP Event Link	Hit Call	Top	Scaled Top	AC50	log AC50	Target Family
CEETOX_H295R_ESTRONE_up	909	-	-	-	ACTIVE	1.24	1.53	1.06	0.0236	steroid hormone
ATG_PXR_TRANS_up	-	NP_071285.1	60	245	ACTIVE	1.89	1.43	21.3	1.33	nuclear receptor
NVS_GPCR_hH1	-	NP_000852.1	-	-	ACTIVE	21.2	1.06	1.42	0.151	gpcr
NVS_GPCR_hAdoRA2a	-	NP_000666.2	-	-	ACTIVE	26.2	1.05	6.95	0.842	gpcr
NVS_GPCR_hAdoRA1	-	NP_000665.1	-	-	ACTIVE	53.5	1.97	6.44	0.809	gpcr

Executive Summary (Beta) | Chemical Properties | Env. Fate/Transport | Hazard | ADME (Beta) | Exposure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

SeqaPASS

United States Environmental Protection Agency

Environmental Topics | Laws & Regulations | About EPA | Search EPA.gov

Related Topics: Safer Chemicals Research

CONTACT US SHARE

Sequence Alignment to Predict Across Species Susceptibility

What is SeqAPASS?

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS), is a fast, online screening tool that allows researchers and regulators to extrapolate toxicity information across species. For some species, such as humans, mice, rats, and zebrafish, the EPA has a large amount of data regarding their toxicological susceptibility to various chemicals. However, the toxicity data for numerous other plants and animals is very limited.

SeqAPASS extrapolates from these data rich model organisms to thousands of other non-target species to evaluate their specific potential

The Assay Table

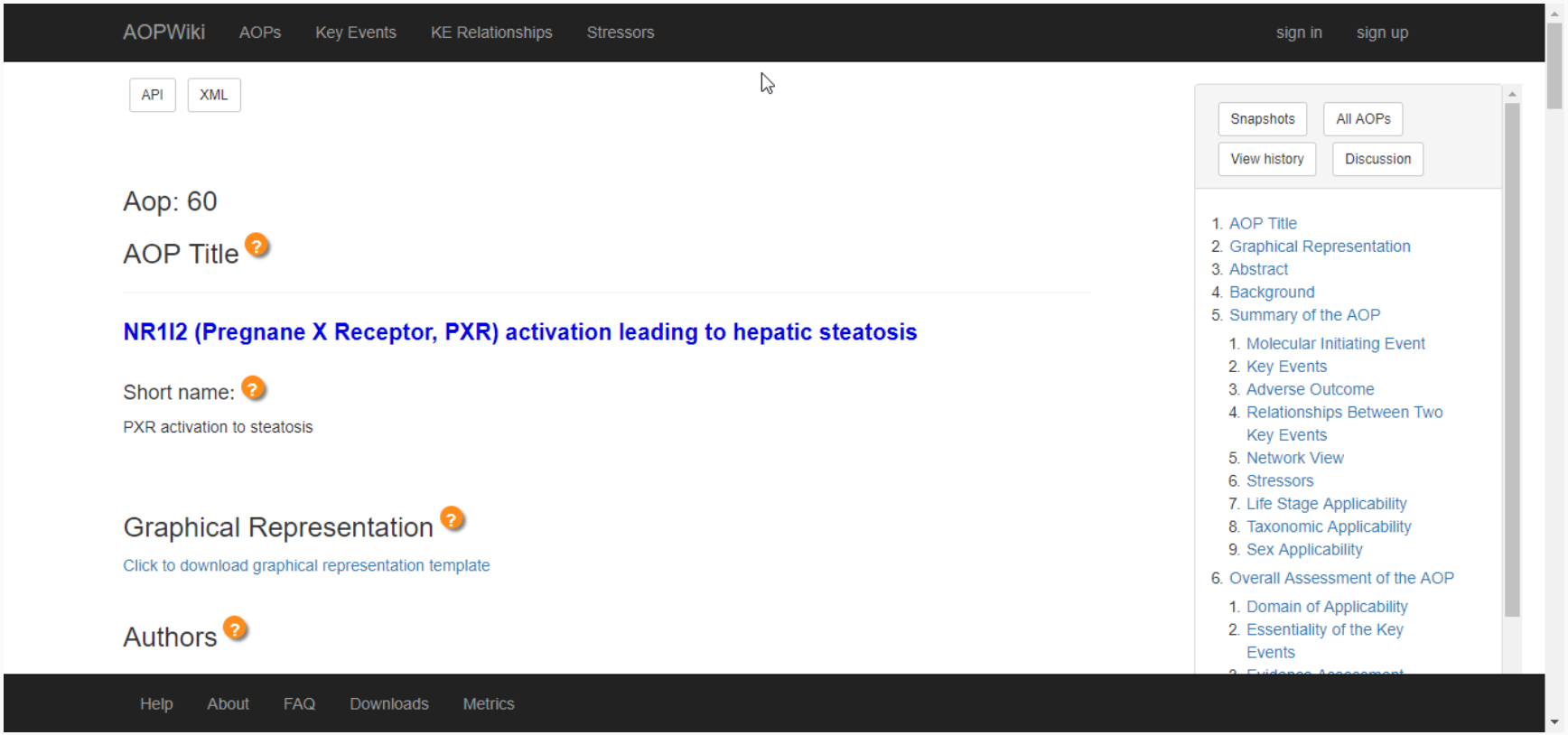
- Links to Adverse Outcome Pathway Wiki

Download as: TSV Excel Show: Inactive Background

Assay Name	Assay De...	SeqAPASS	AOP Link †	AOP Event Link	Hit Call	Top	Scaled Top	AC50	log AC50	Target Family
CEETOX_H295R_ESTRONE_up	909	-	-	-	ACTIVE	1.24	1.53	1.06	0.0236	steroid hormone
ATG_PXR_TRANS_up	-	NP_071285.1	60	245	ACTIVE	1.89	1.43	21.3	1.33	nuclear receptor
NVS_GPCR_hH1	-	NP_000852.1	-	-	ACTIVE	21.2	1.06	1.42	0.151	gpcr
NVS_GPCR_hAdoRA2a	-	NP_000666.2	-	-	ACTIVE	26.2	1.05	6.95	0.842	gpcr
NVS_GPCR_hAdoRA1	-	NP_000665.1	-	-	ACTIVE	53.5	1.97	6.44	0.809	gpcr

Executive Summary (Beta) | Chemical Properties | Env. Fate/Transport | Hazard | ADME (Beta) | Exposure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

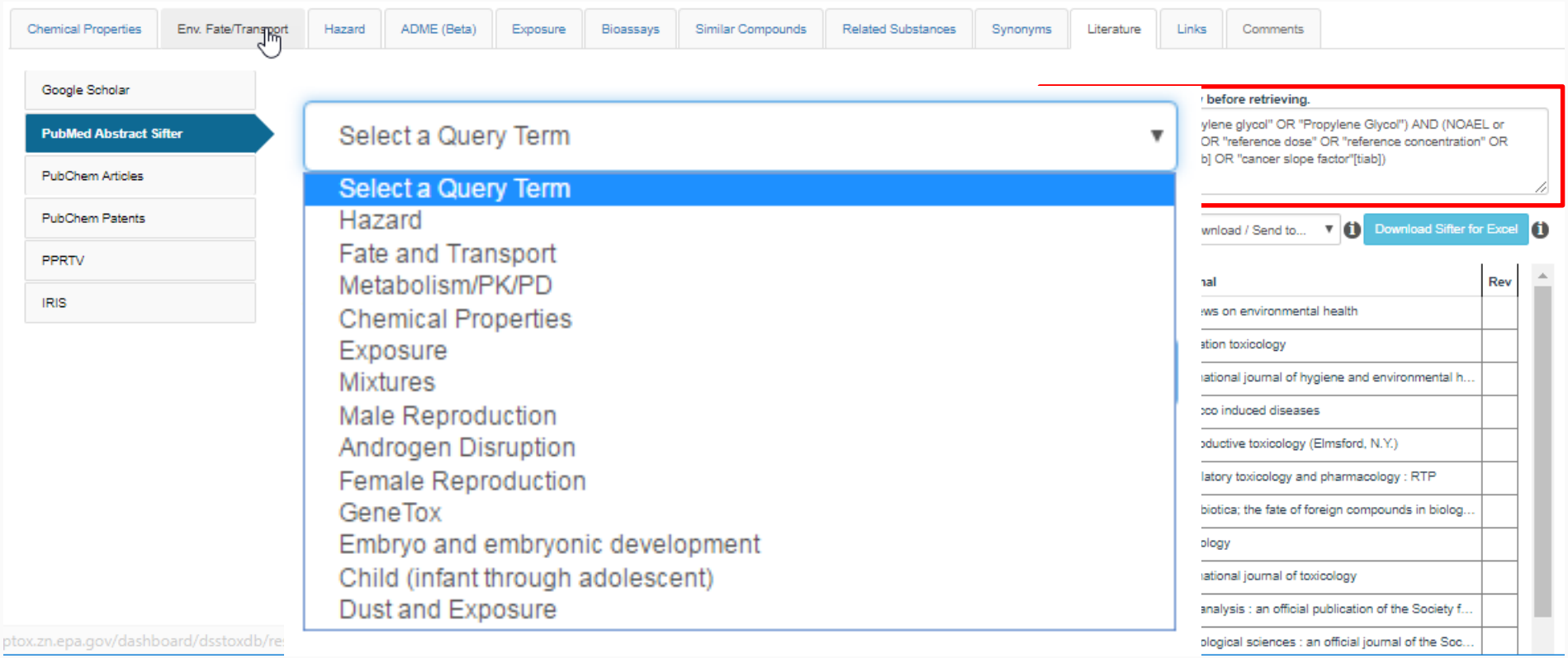
AOPWiki Integration



Single Click Download

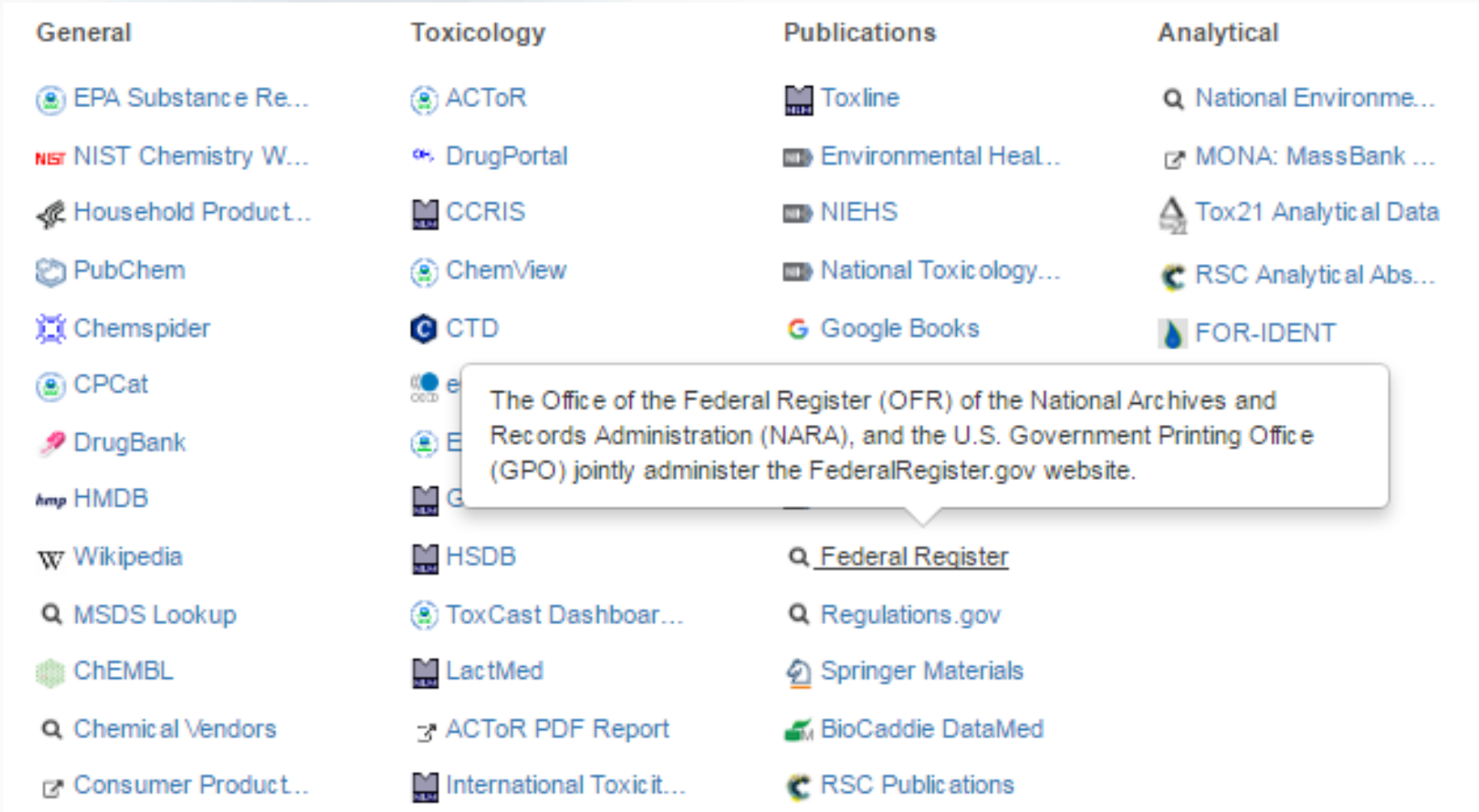
	A	B	C	D	E	F	G	H	I	J	K	L
1	**Assay Name**	**Assay Description**	**SeqAPASS**	**AOP Link**	**AOP Event Link**	**Hit Call**	**Top**	**Scaled To**	**AC50**	**logAC50**	**Intended Target Family**	
2	ATG_TCF_b_cat_CIS_dn	-	-	-	-	ACTIVE	0.872	1.27	0.869	-0.0609	dna binding	
3	NVS_NR_hFXR_Antagonist	716	NP_001193922.1	61	479	ACTIVE	35.6	1.41	0.165	-0.783	nuclear receptor	
4	ATG_XTT_Cytotoxicity_up	-	-	-	-	ACTIVE	98.7	1.21	0.00800	-2.10	cell cycle	
5	ATG_Oct_MLP_CIS_dn	-	-	-	-	ACTIVE	0.838	1.19	0.0936	-1.03	dna binding	
6	ATG_AP_2_CIS_dn	-	-	-	-	ACTIVE	0.806	2.56	0.101	-0.997	dna binding	
7	ATG_M_19_TRANS_up	-	-	-	-	ACTIVE	0.608	1.67	0.000901	-3.05	background measurement	
8	ATG_M_06_TRANS_dn	-	-	-	-	ACTIVE	0.349	1.33	0.0101	-2.00	background measurement	
9	ATG_GLI_CIS_dn	-	-	-	-	ACTIVE	0.905	2.32	0.443	-0.354	dna binding	
10	ATG_M_61_TRANS_dn	-	-	-	-	ACTIVE	0.349	1.33	0.0101	-2.00	background measurement	
11	ATG_HIF1a_CIS_dn	-	-	-	-	ACTIVE	1.41	1.80	0.937	-0.0282	dna binding	
12	ATG_HNF6_CIS_dn	-	-	-	-	ACTIVE	0.386	1.07	0.643	-0.192	dna binding	
13	TOX21_AR_BLA_Agonist_ch1	-	-	-	-	INACTIVE	0.00	0.00	-	-	background measurement	
14	TOX21_AR_BLA_Agonist_ch2	-	-	-	-	INACTIVE	0.00	0.00	-	-	background measurement	
15	TOX21_AR_BLA_Agonist_ratio	761	P10275.2	187	1134	INACTIVE	0.00	0.00	-	-	nuclear receptor	
16	TOX21_AR_BLA_Antagonist_ratio	762	P10275.2	187	1134	INACTIVE	0.00	0.00	-	-	nuclear receptor	
17	TOX21_AR_BLA_Antagonist_viability	-	-	-	-	INACTIVE	0.00	0.00	-	-	cell cycle	
18	TOX21_AR_LUC_MDAKB2_Agonist	764	P10275.2	187	1134	INACTIVE	0.00	0.00	-	-	nuclear receptor	
19	TOX21_AR_LUC_MDAKB2_Antagonist	765	P10275.2	187	1134	INACTIVE	0.00	0.00	-	-	nuclear receptor	
20	TOX21_Aromatase_Inhibition	-	NP_000094.2	153	964	INACTIVE	0.00	0.00	-	-	cyp	
21	TOX21_AutoFluor_HEK293_Cell_blue	-	-	-	-	INACTIVE	0.00	0.00	-	-	background measurement	

Literature Searches and Links
Rebuilt Abstract Sifter

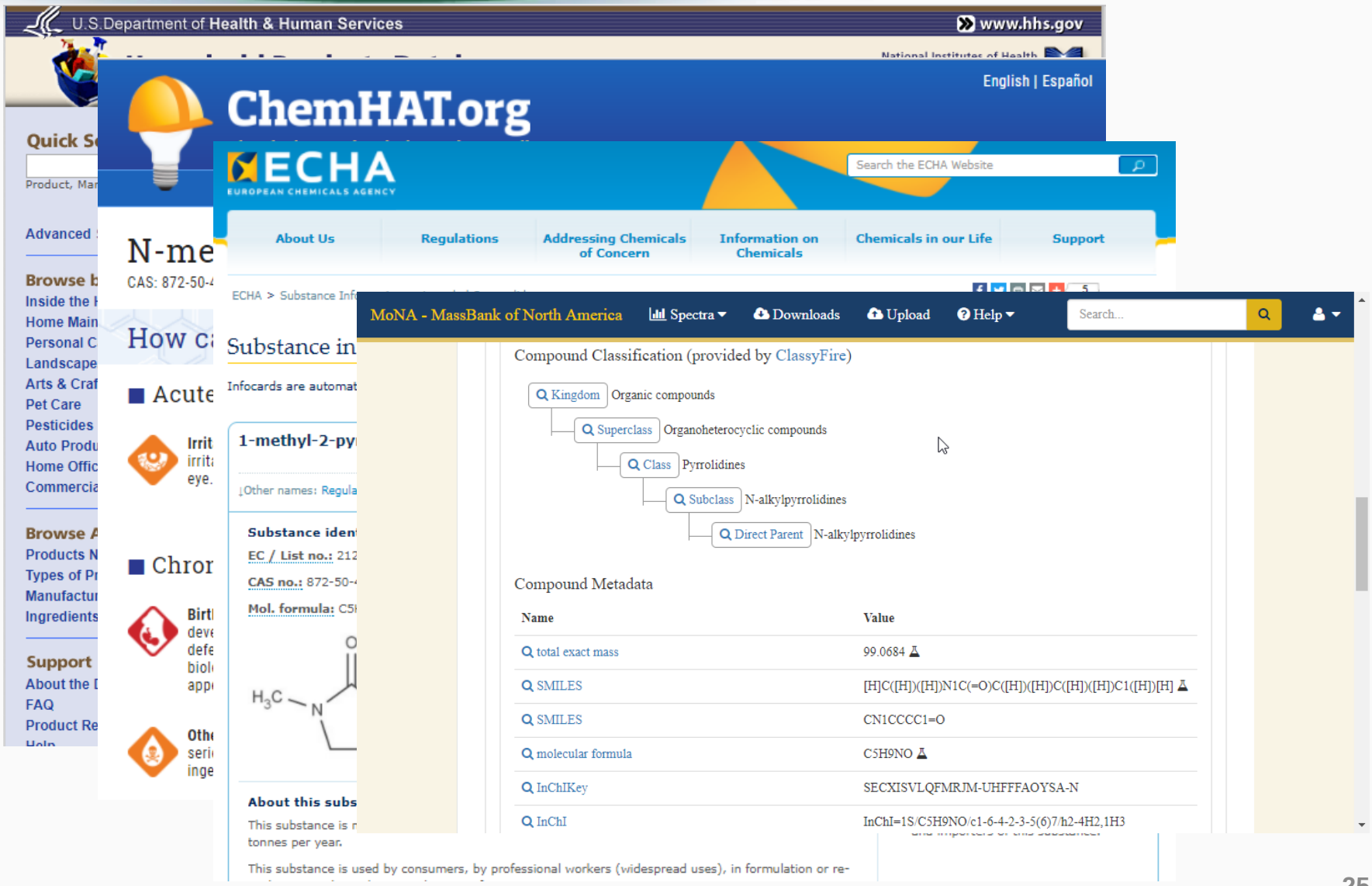


Executive Summary (Beta)
Chemical Properties
Env. Fate/Transport
Hazard
ADME (Beta)
Exposure
Bioassays
Similar Compounds
Related Substances
Synonyms
Literature
Links
Comments

Links to Other Resources



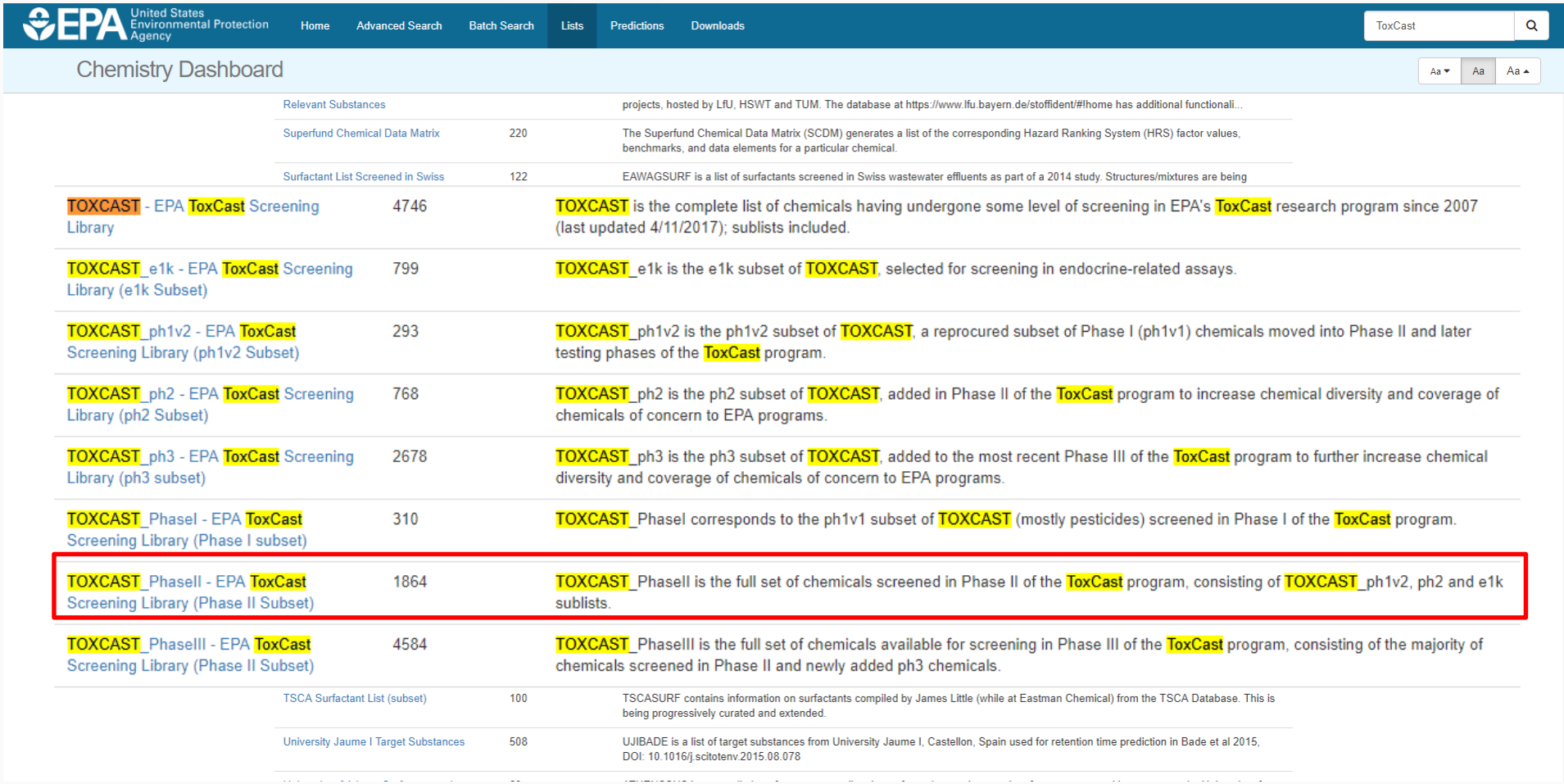
Example External Links…



Accessing Lists of Chemicals

- Build out definitive “lists” of chemicals
  - Algal toxins
  - Pesticides
  - **Toxcast screening chemical collection**

The Collection of Lists



ToxCast phases over time

This is an open access article published under a Creative Commons Attribution (CC-BY) License, which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

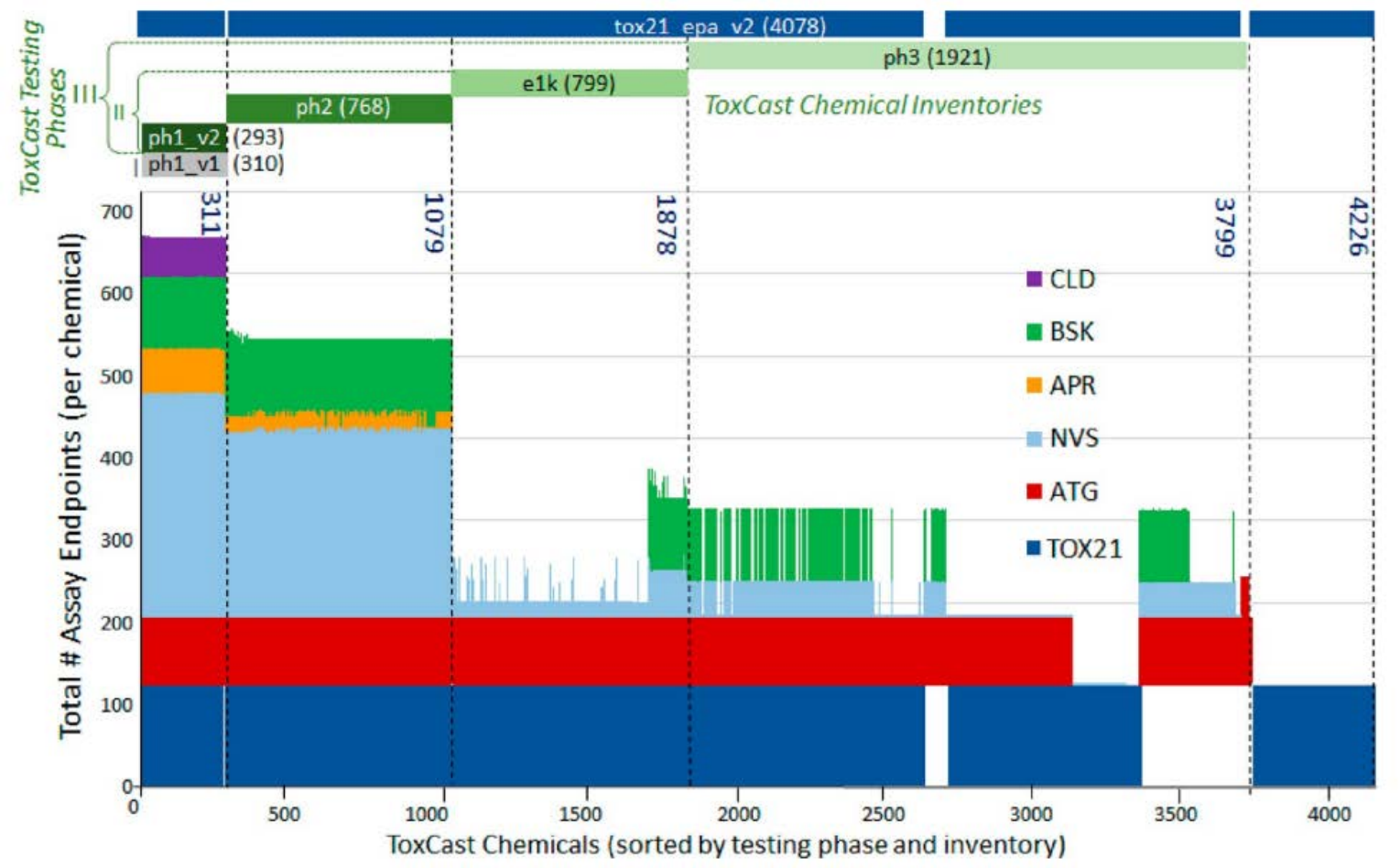
Chemical Research in Toxicol

Perspective

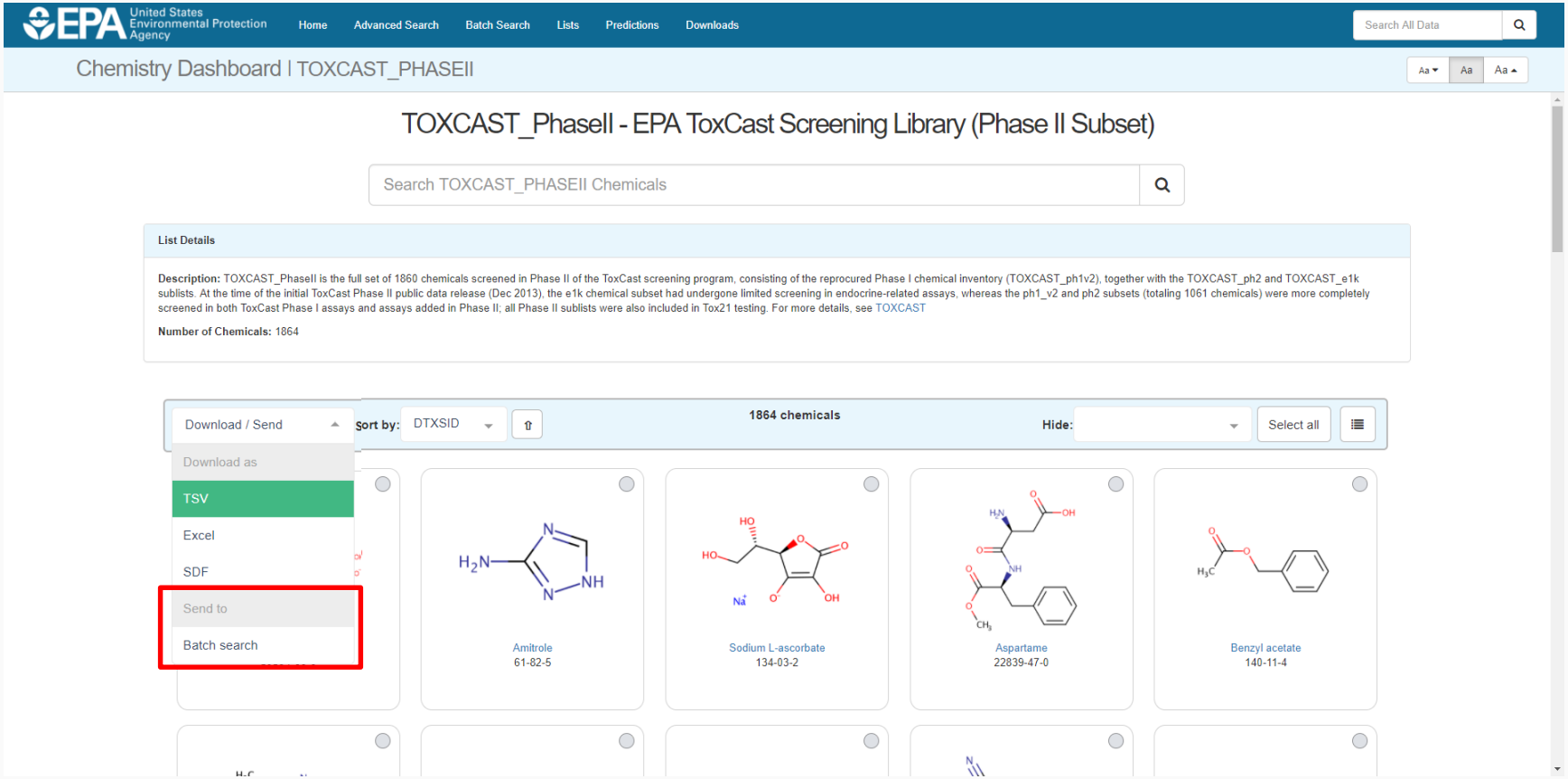
pubs.acs.org/crt

ToxCast Chemi
Toxicology

Ann M. Richard,*,† Ri
Inthirany Thillainadara
John F. Wambaugh,† T
Antony J. Williams,† S



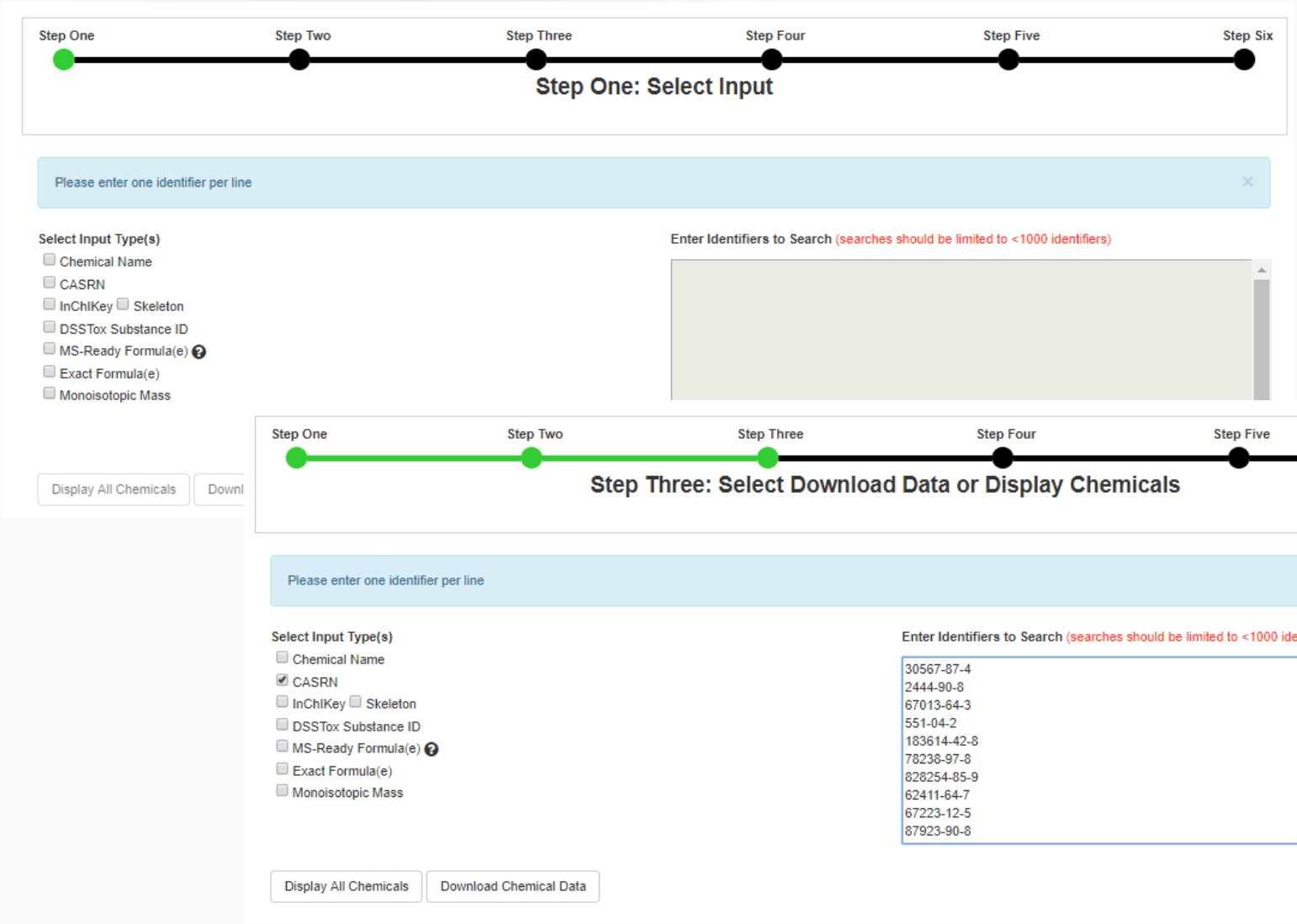
Single Click to View Data



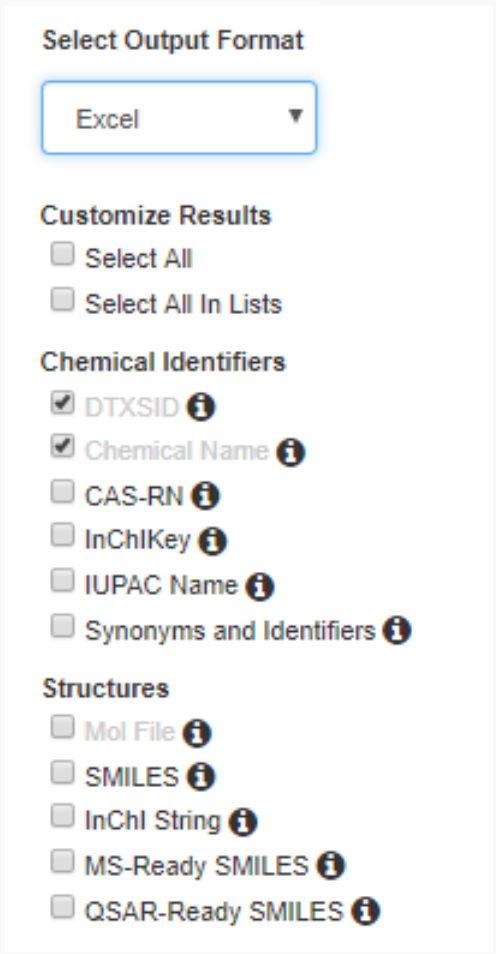
Answering Questions

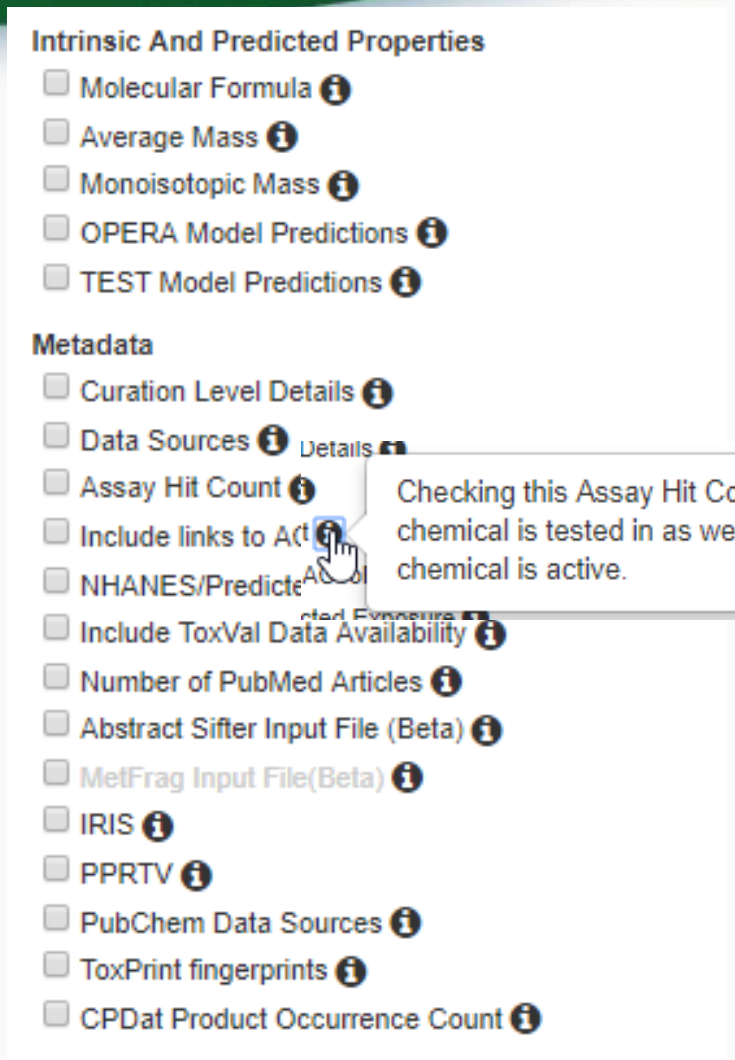
- I have a 1000 CAS Numbers (or Names) – are there data available?
  - Has any Toxcast data been run?
  - Are there Toxicity Data values available?
  - Are there predicted exposure data (via Expocast)?
  - Can I get predicted physchem data for my model?

Batch Searching for Data for Thousands of Chemicals

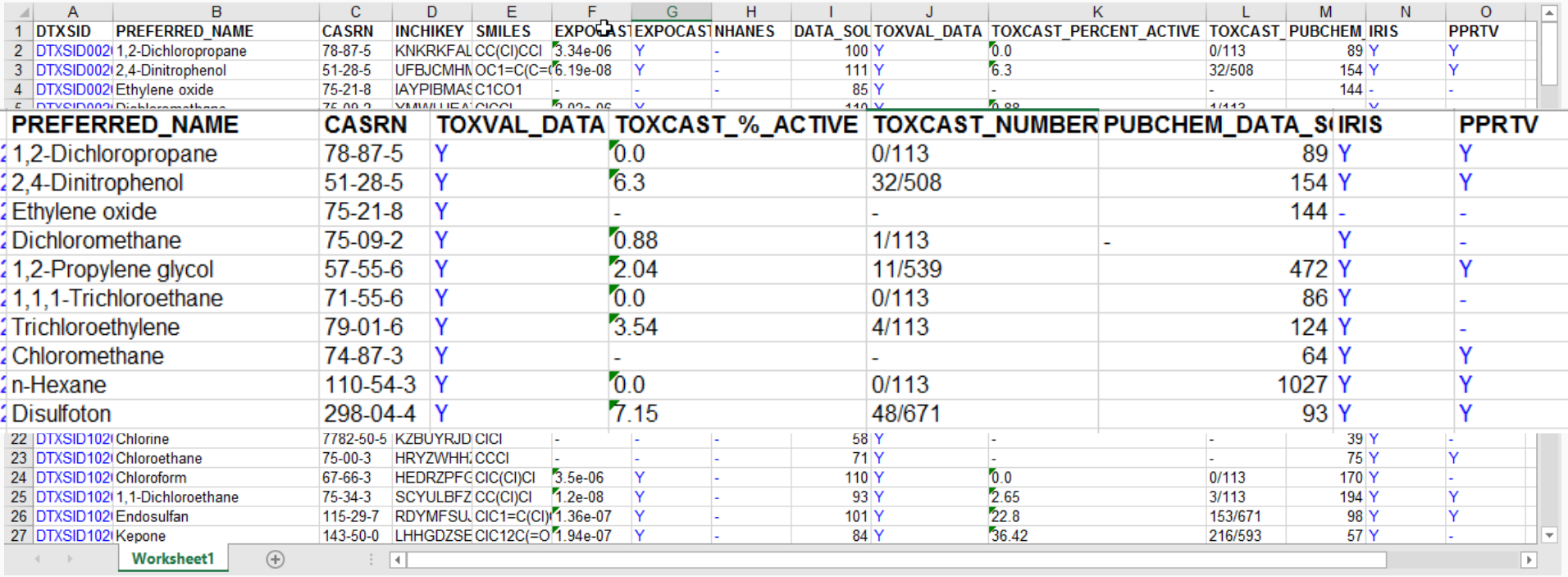


Batch Search





Downloadable Results



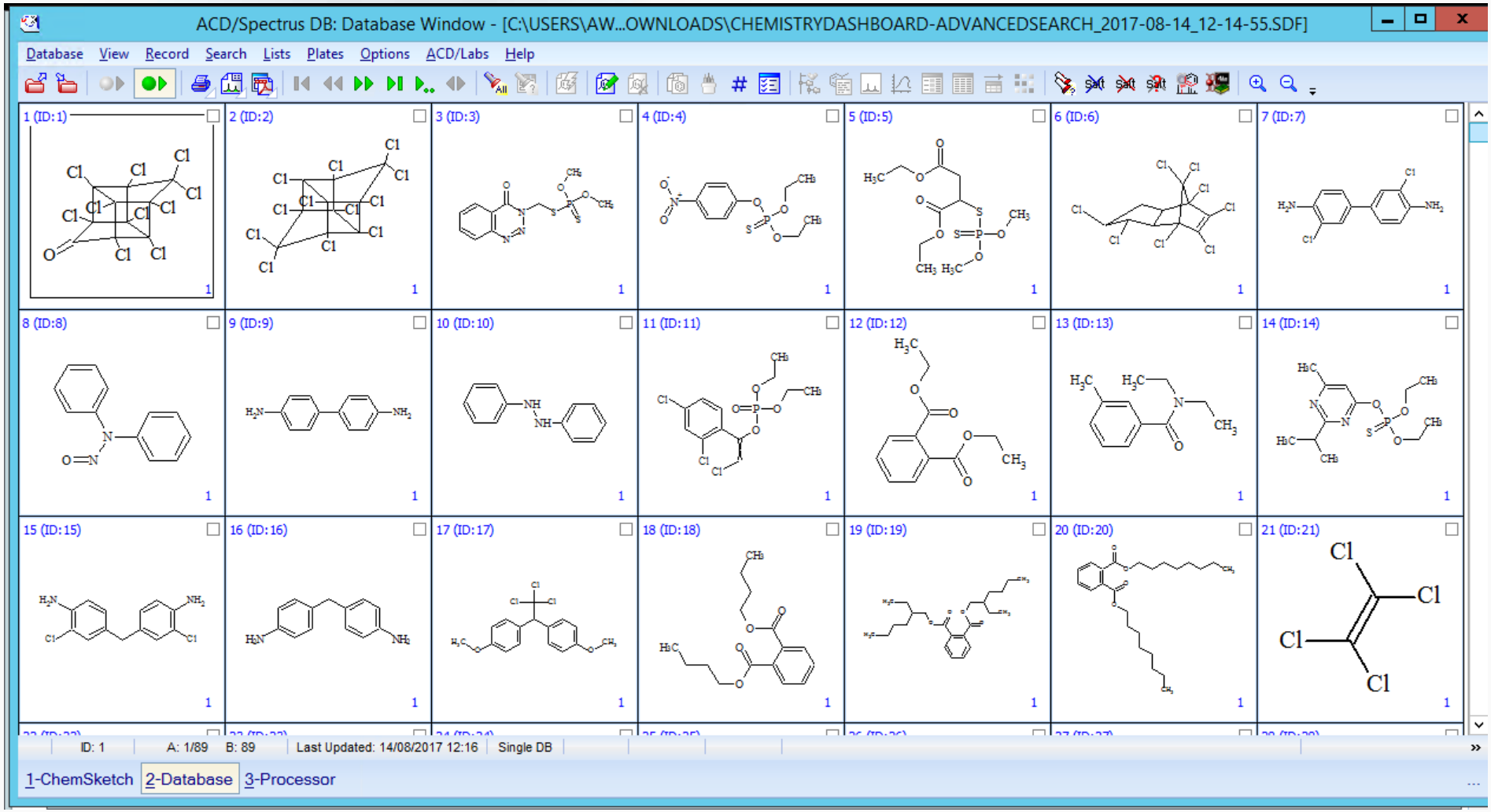
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	DTXSID	PREFERRED_NAME	CASRN	INCHIKEY	SMILES	EXPOCAST	EXPOCAST	NHANES	DATA_SOU	TOXVAL_DATA	TOXCAST_PERCENT_ACTIVE	TOXCAST_	PUBCHEM_	IRIS	PPRTV
2	DTXSID002	1,2-Dichloropropane	78-87-5	KNKRKFAL	CC(Cl)CCl	3.34e-06	Y	-	100	Y	0.0	0/113	89	Y	Y
3	DTXSID002	2,4-Dinitrophenol	51-28-5	UFBJCMHN	OC1=C(C=(	6.19e-08	Y	-	111	Y	6.3	32/508	154	Y	Y
4	DTXSID002	Ethylene oxide	75-21-8	IAYPIBMAS	C1CO1	-	-	-	85	Y	-	-	144	-	-
5	DTXSID002	Dichloromethane	75-09-2	YMWUUEA	ClCCl	2.02e-06	Y		110	Y	0.88	1/113		Y	

	PREFERRED_NAME	CASRN	TOXVAL_DATA	TOXCAST_%_ACTIVE	TOXCAST_NUMBER	PUBCHEM_DATA_S	IRIS	PPRTV
2	1,2-Dichloropropane	78-87-5	Y	0.0	0/113	89	Y	Y
3	2,4-Dinitrophenol	51-28-5	Y	6.3	32/508	154	Y	Y
4	Ethylene oxide	75-21-8	Y	-	-	144	-	-
5	Dichloromethane	75-09-2	Y	0.88	1/113	-	Y	-
6	1,2-Propylene glycol	57-55-6	Y	2.04	11/539	472	Y	Y
7	1,1,1-Trichloroethane	71-55-6	Y	0.0	0/113	86	Y	-
8	Trichloroethylene	79-01-6	Y	3.54	4/113	124	Y	-
9	Chloromethane	74-87-3	Y	-	-	64	Y	Y
10	n-Hexane	110-54-3	Y	0.0	0/113	1027	Y	Y
11	Disulfoton	298-04-4	Y	7.15	48/671	93	Y	Y

22	DTXSID102	Chlorine	7782-50-5	KZBUYRJD	ClCl	-	-	-	58	Y	-	-	39	Y	-
23	DTXSID102	Chloroethane	75-00-3	HRYZWHH	CCCl	-	-	-	71	Y	-	-	75	Y	Y
24	DTXSID102	Chloroform	67-66-3	HEDRZPFC	ClC(Cl)Cl	3.5e-06	Y	-	110	Y	0.0	0/113	170	Y	-
25	DTXSID102	1,1-Dichloroethane	75-34-3	SCYULBFZ	CC(Cl)Cl	1.2e-08	Y	-	93	Y	2.65	3/113	194	Y	Y
26	DTXSID102	Endosulfan	115-29-7	RDYMFSU	ClC1=C(Cl)(	1.36e-07	Y	-	101	Y	22.8	153/671	98	Y	Y
27	DTXSID102	Kepone	143-50-0	LHHGDZSE	ClC12C(=O	1.94e-07	Y	-	84	Y	36.42	216/593	57	Y	-

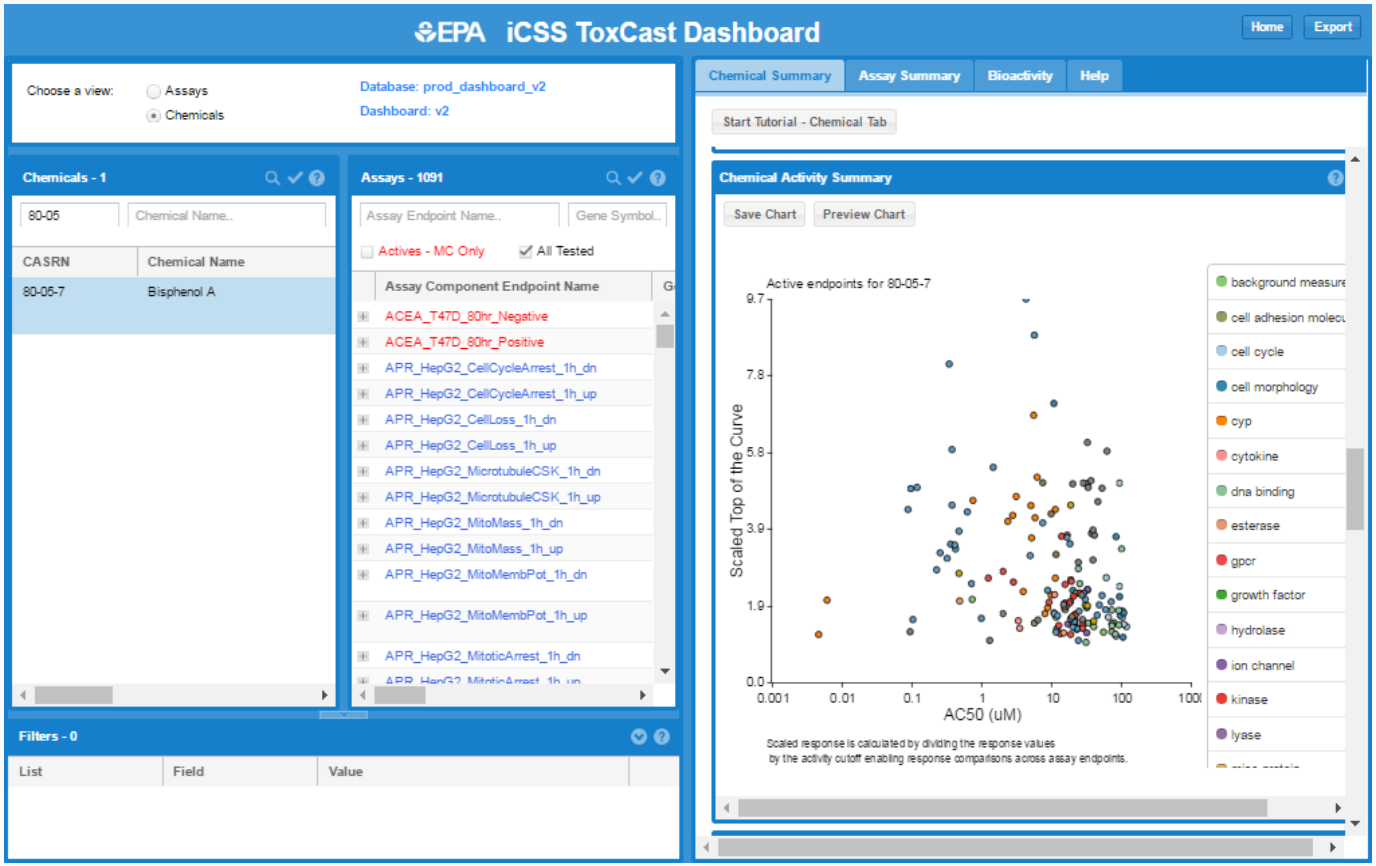
Worksheet1

SDF Download

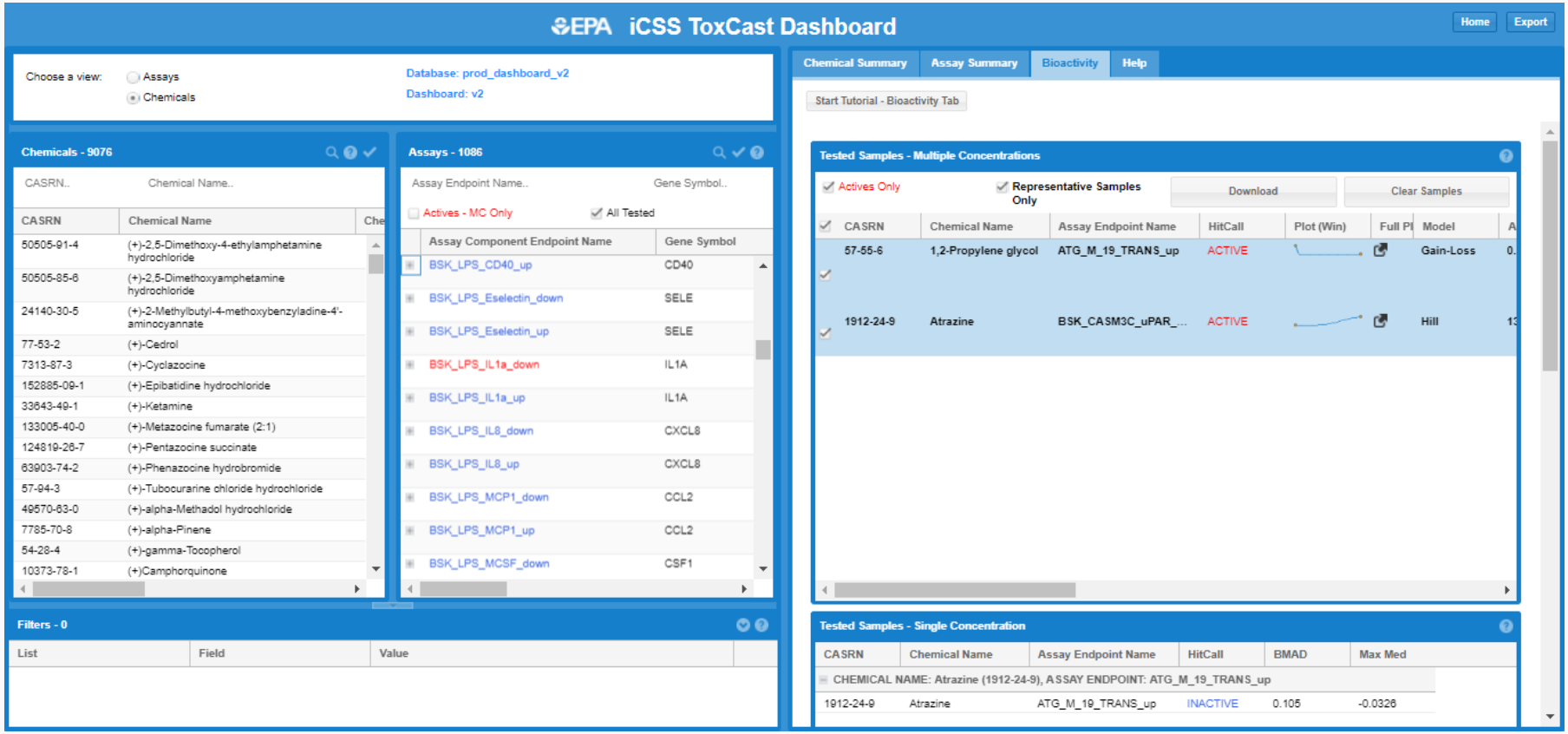


INCOMPLETE support for *in vitro* HTS data – use ToxCast Dashboard

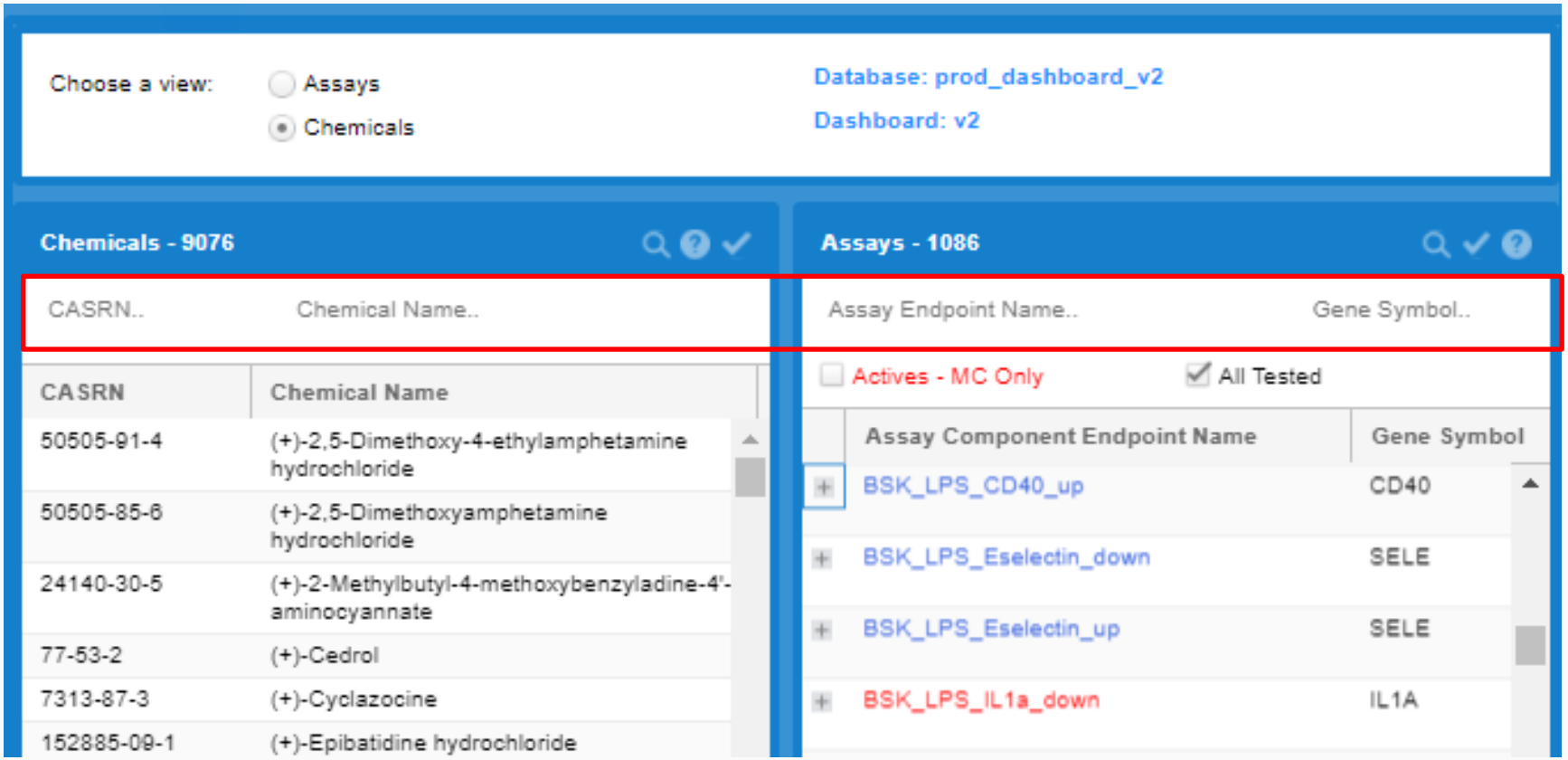
- Data from ToxCast and the Tox21 collaboration
- https://actor.epa.gov/dashboard/



ToxCast Dashboard

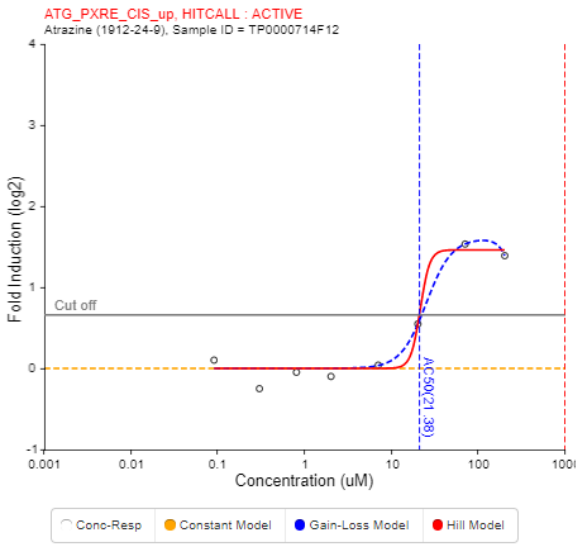


ToxCast Dashboard



ToxCast Dashboard

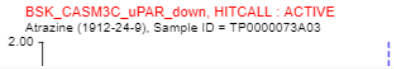
Save Chart | Preview Chart



Model	AIC	RMSE	Top	AC50	Slope
Constant	19.75	0.76	-	-	-
Gain-Loss	-2.96	0.1	1.61	25.12	2.97
Hill	-4.72	0.11	1.46	21.38	8

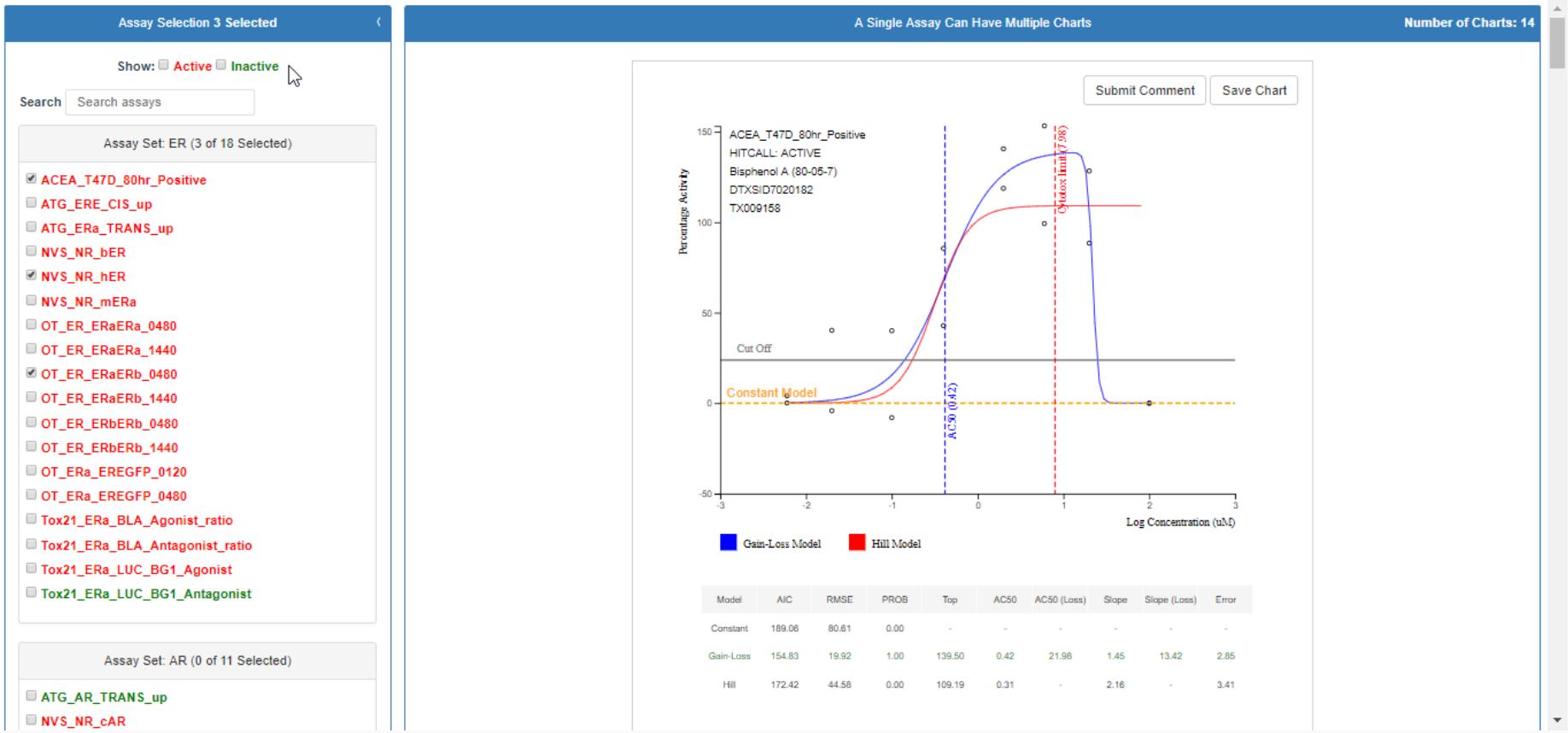
Flag
none

Save Chart | Preview Chart

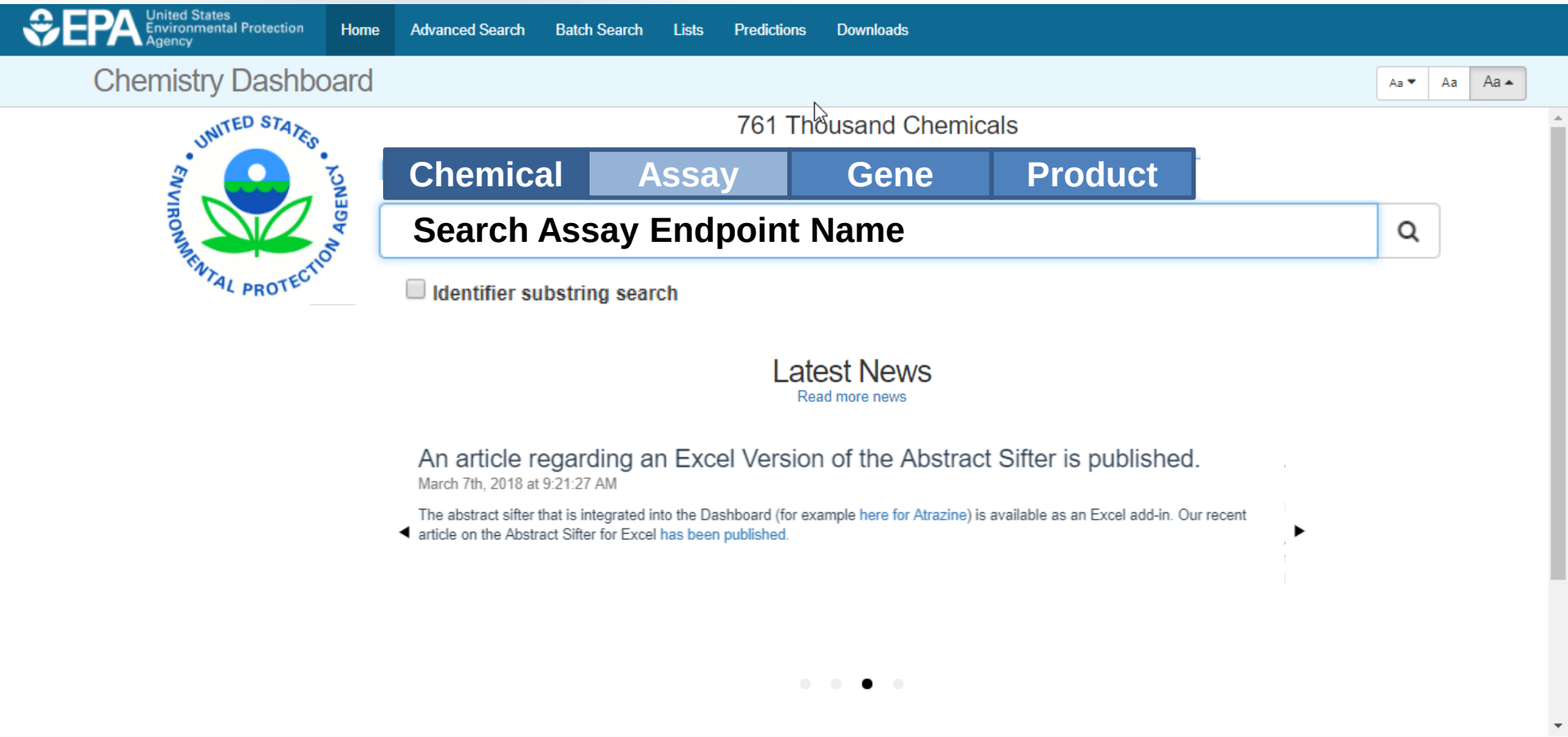


Model	AIC	RMSE	Top	AC50	Slope
Constant	-6.51	0.08	-	-	-

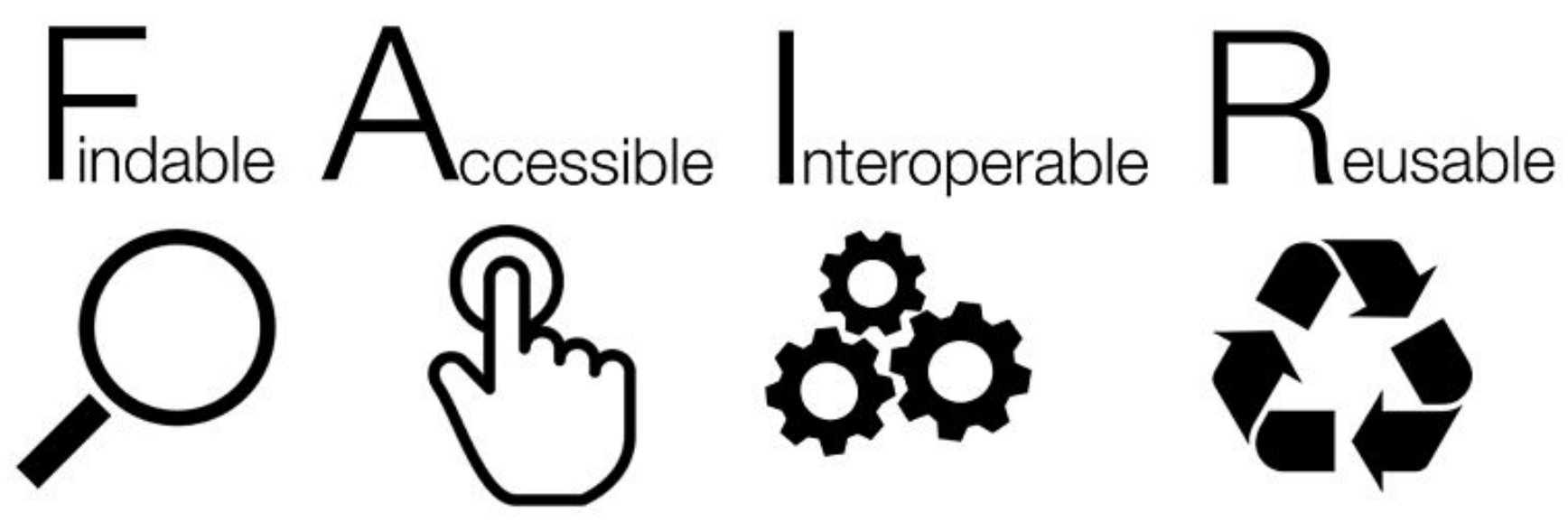
Work in Progress – Bioactivity Curves



Future Search

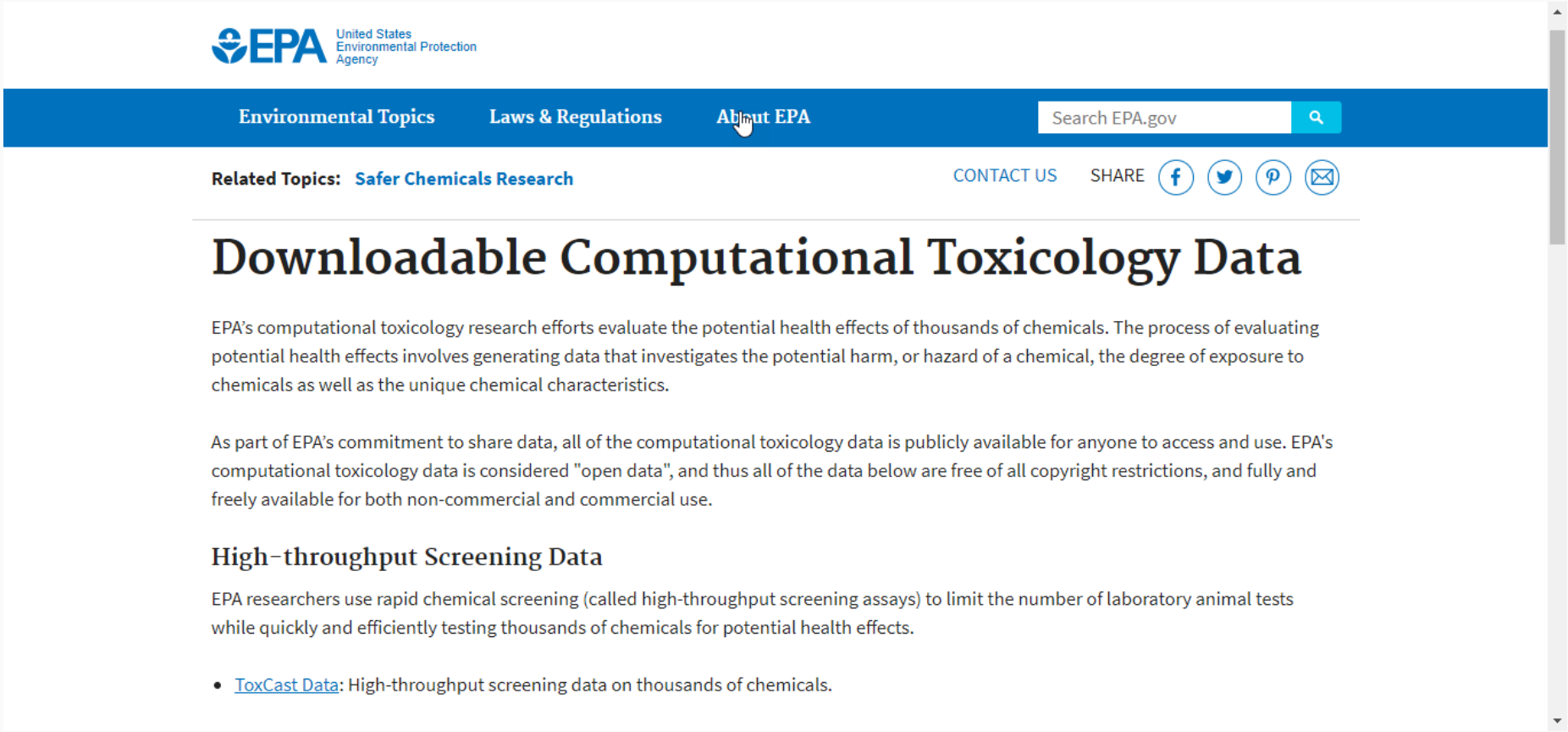


Our support for FAIR Data

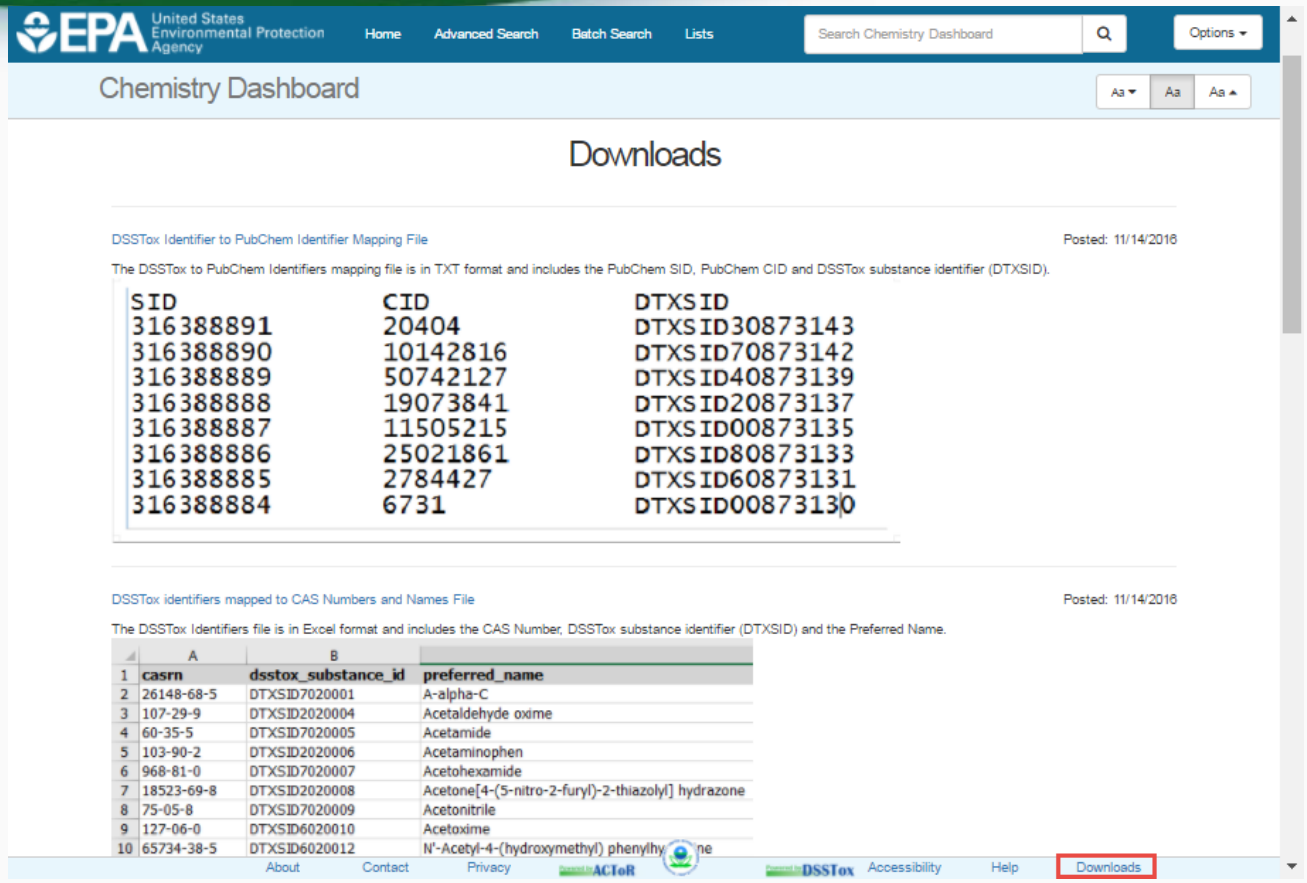


Downloadable ToxCast Data

https://www.epa.gov/chemical-research/downloadable-computational-toxicology-data

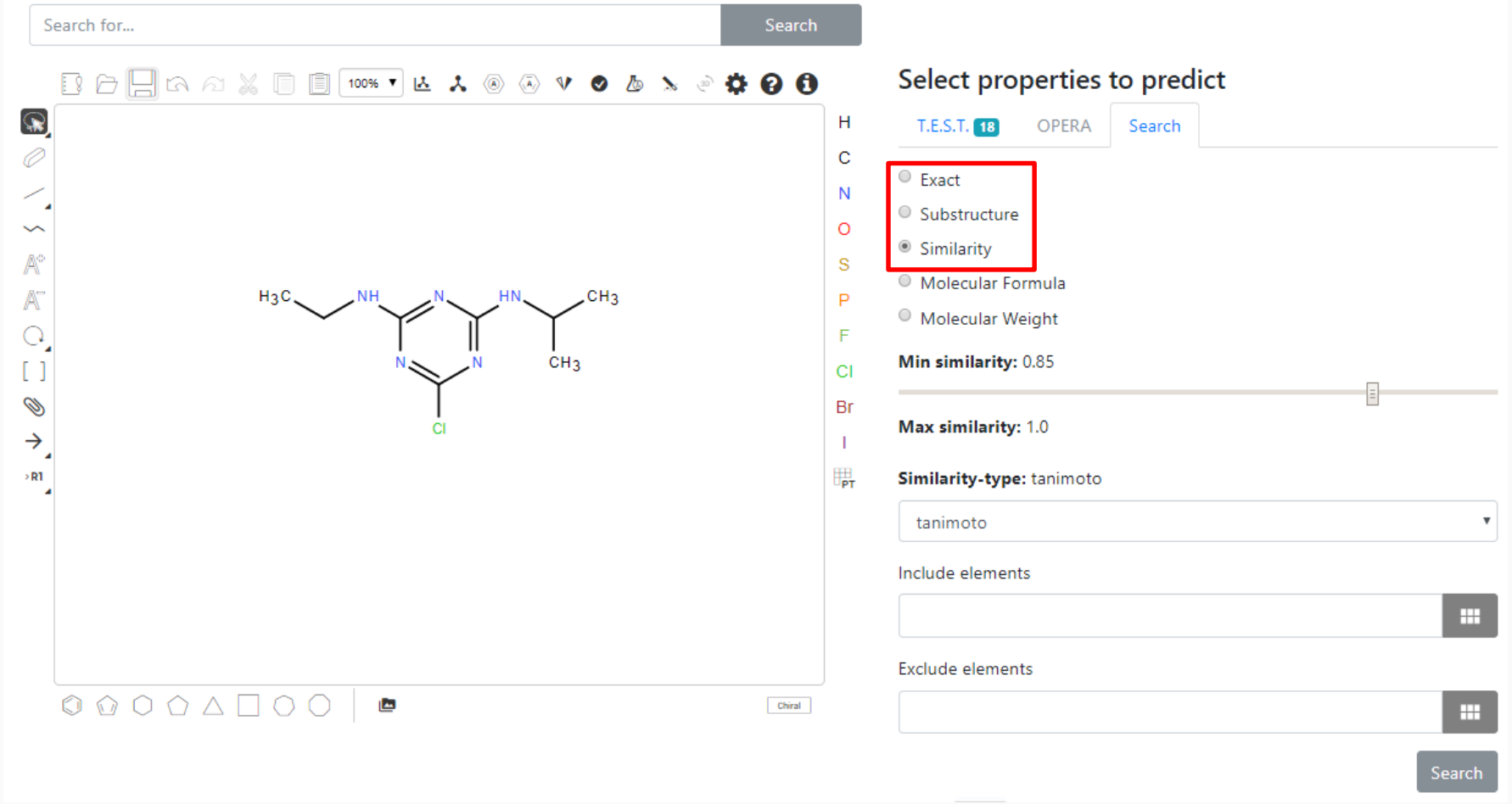


Delivering our Chemistry Data

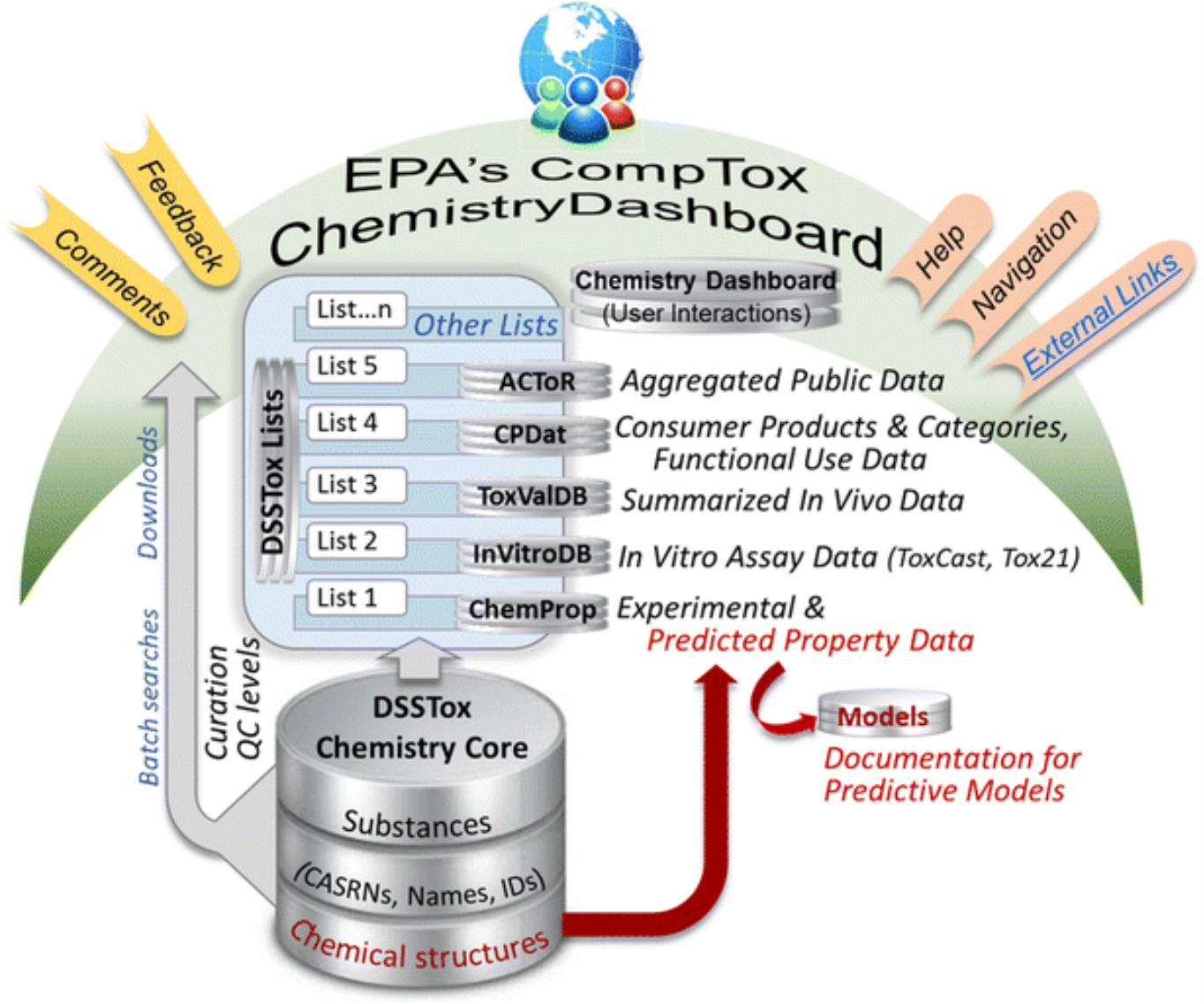


- Various types of data at FTP download site: ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_Data/Chemistry_Dashboard

When you have a structure editor…



What we are building…



Conclusion

- The CompTox Chemistry Dashboard provides access to data for ~760,000 chemicals
- High quality data from ongoing curation efforts
- An integration hub for multiple “modules”
  - Experimental and predicted properties
  - Human and Ecological Hazard data
  - Exposure data – products, data in the environment
  - *In vitro* bioassay data – ToxCast/Tox21
  - Literature searching – Google Scholar and PubMed
  - Specialized searches – mass/formula for analytical support
  - Batch searching and Real Time Predictions
- The primary architecture for NCCT data

Acknowledgements

EPA-RTP

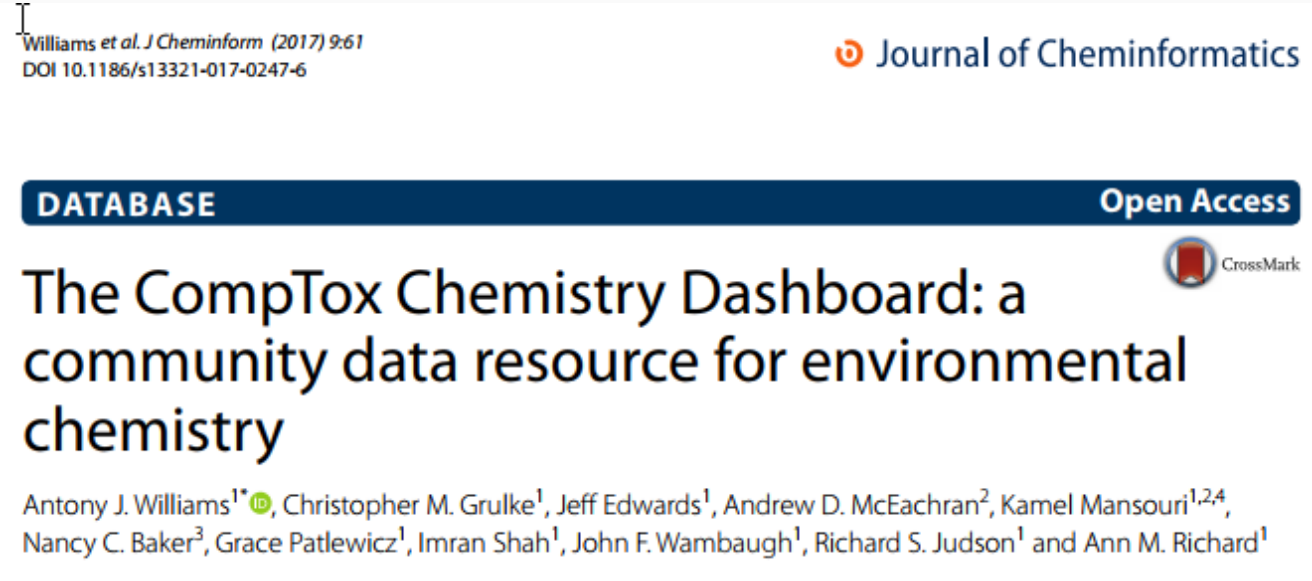
The NCCT Software Development Group

An enormous team of contributors from NCCT

and collaborators from
NERL
NHERL
NRMRL

Contact

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



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