CompTox Chemicals Dashboard v3 and invitroDB v3.1

http://comptox.epa.gov/dashboard

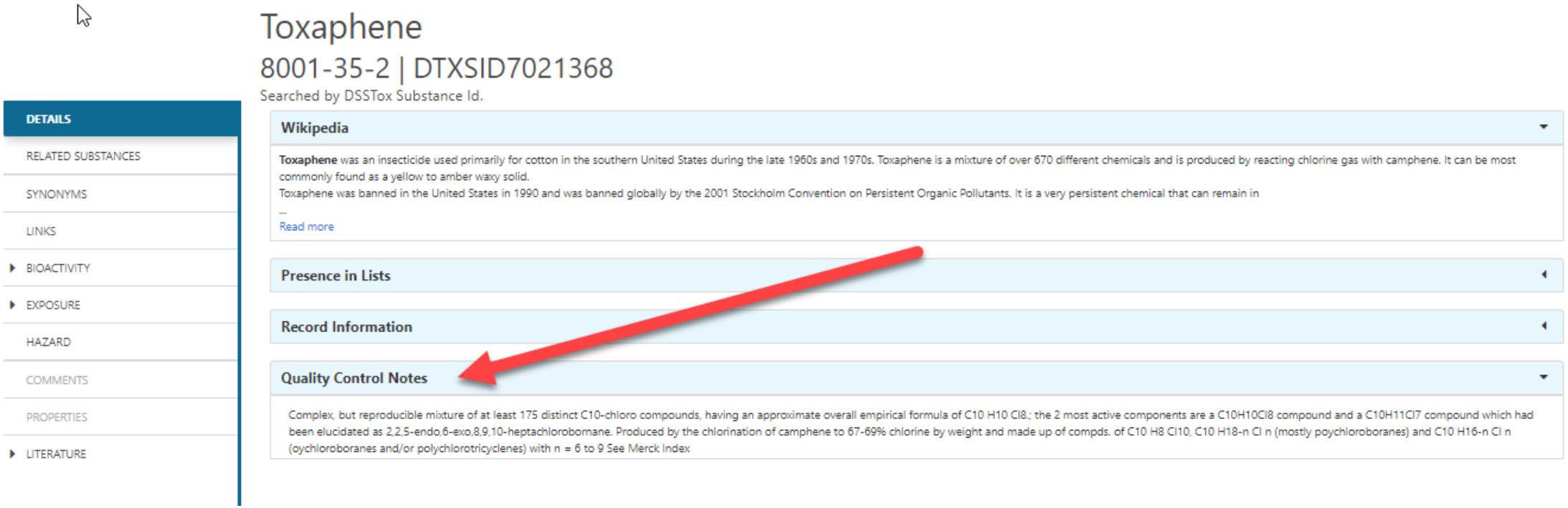
Katie Paul Friedman and Antony Williams
National Center for Computational Toxicology

Overview of changes since August 2018 release

- Data
  - 875k chemicals total. An additional 110k chemical substances added
  - *InvitroDBv3.1* – including updated assay descriptions
  - ToxVal v7 data integrated – includes enormous curation effort
  - New OPERA predictions

- New User Interface elements
  - Reworked tables across the application
  - Reworked multiple chemical results page
  - Navigating concentration response plots for all AEIDs in *invitroDB_v3* data, not just the EDSP21 assays
  - Enhanced batch search capabilities

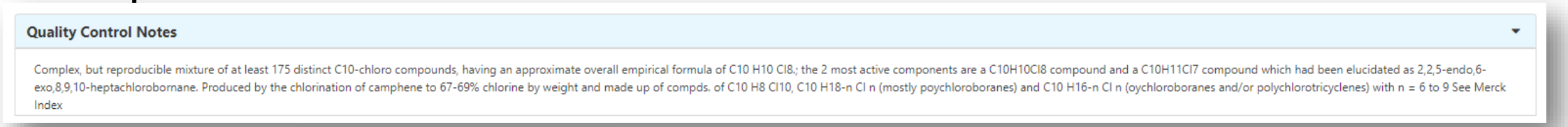
QC Notes for Chemicals



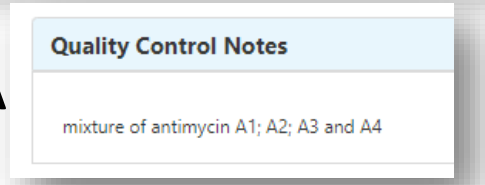
Accessing QC Notes for Chemicals

Examples

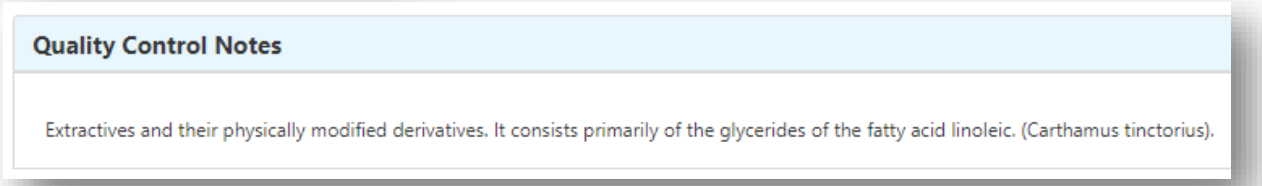
- Toxaphene



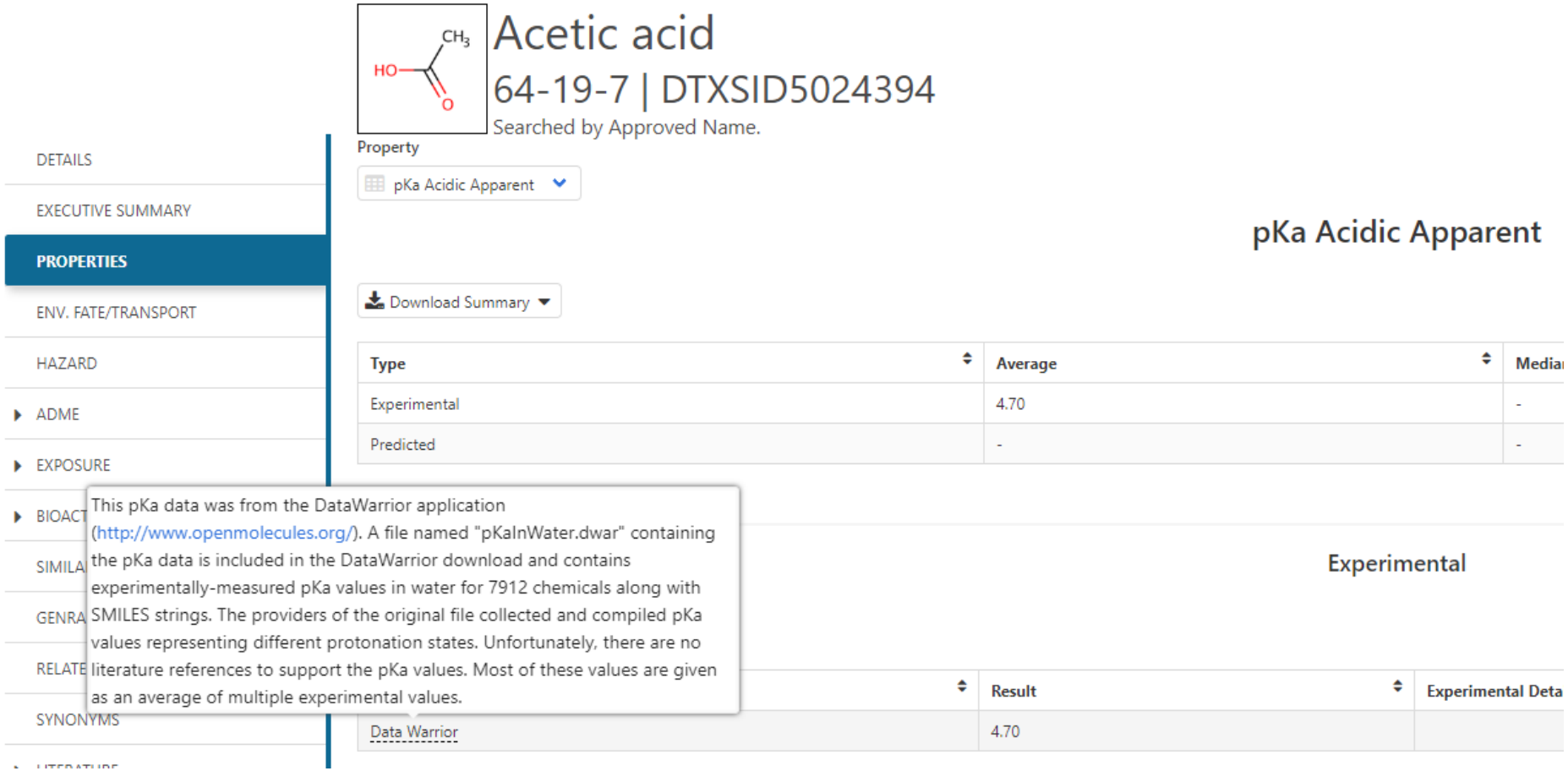
- Antimycin A



- Safflower Oil

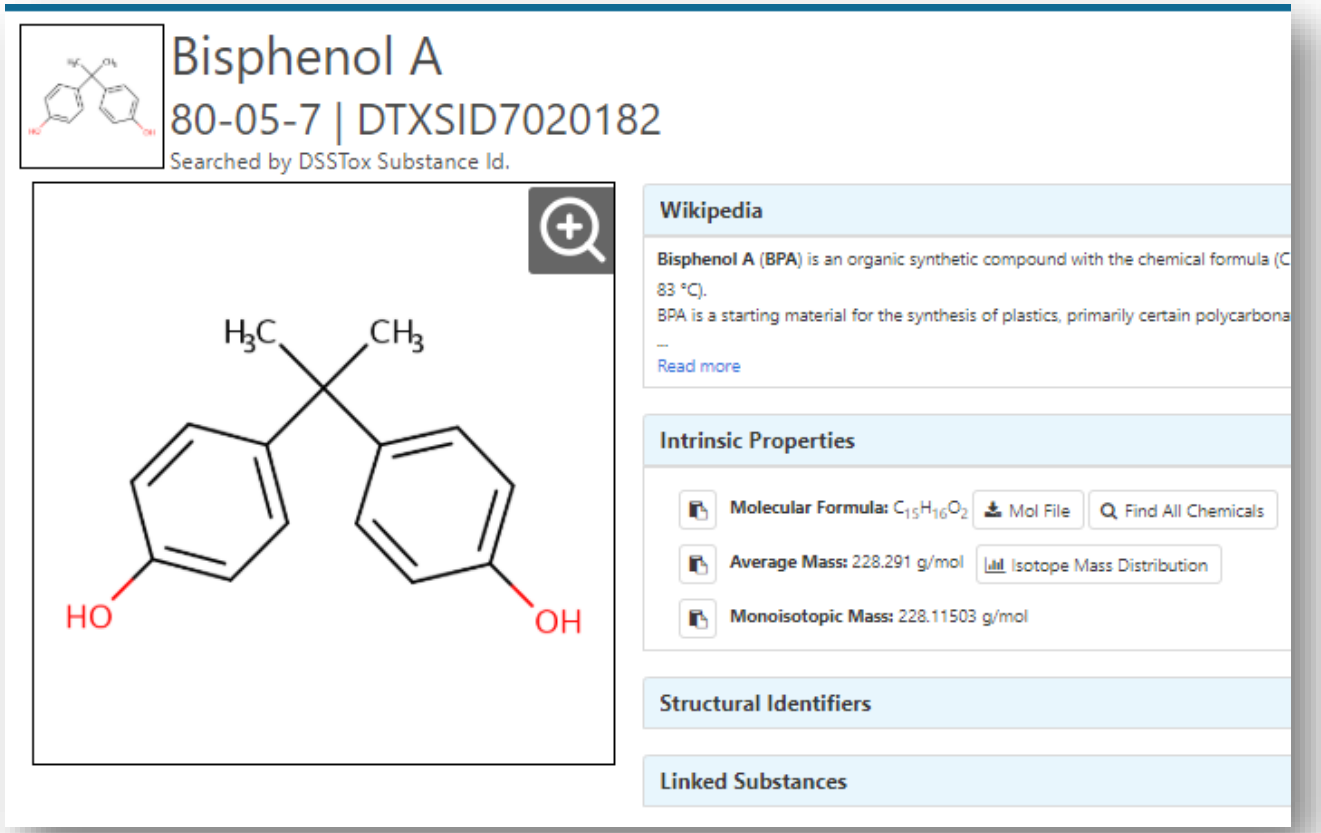


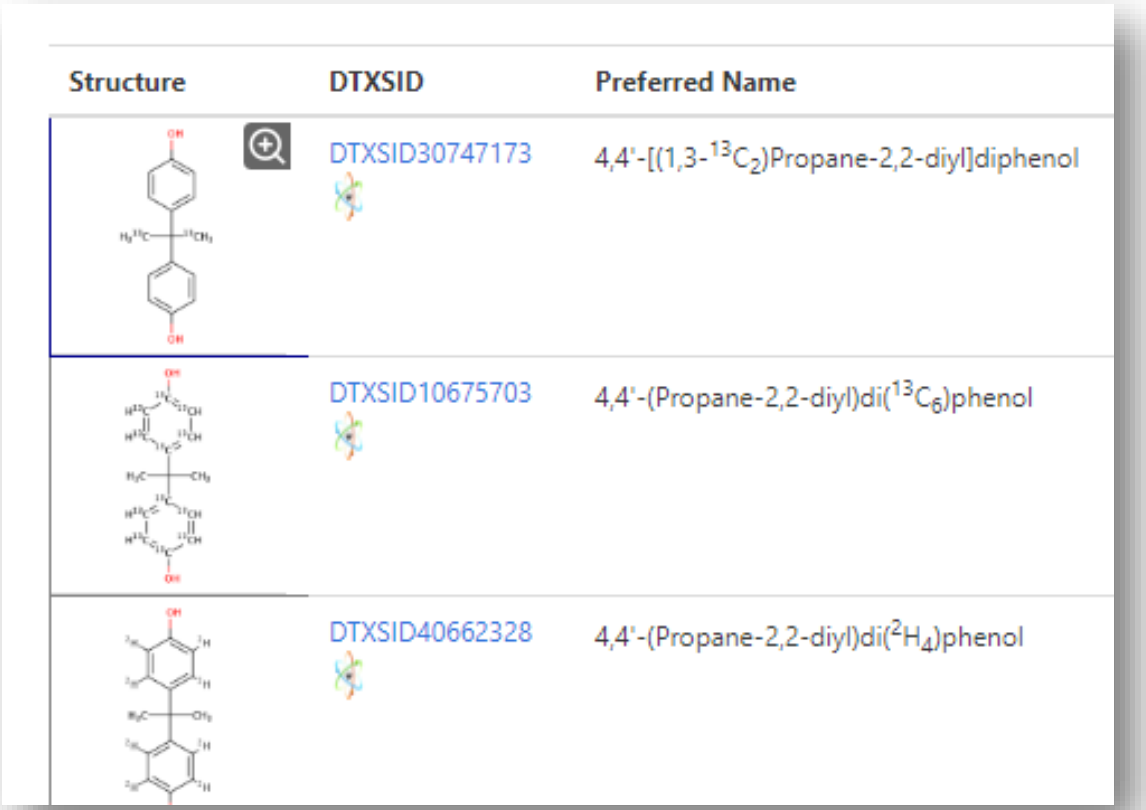
pKa experimental data added – no predictions



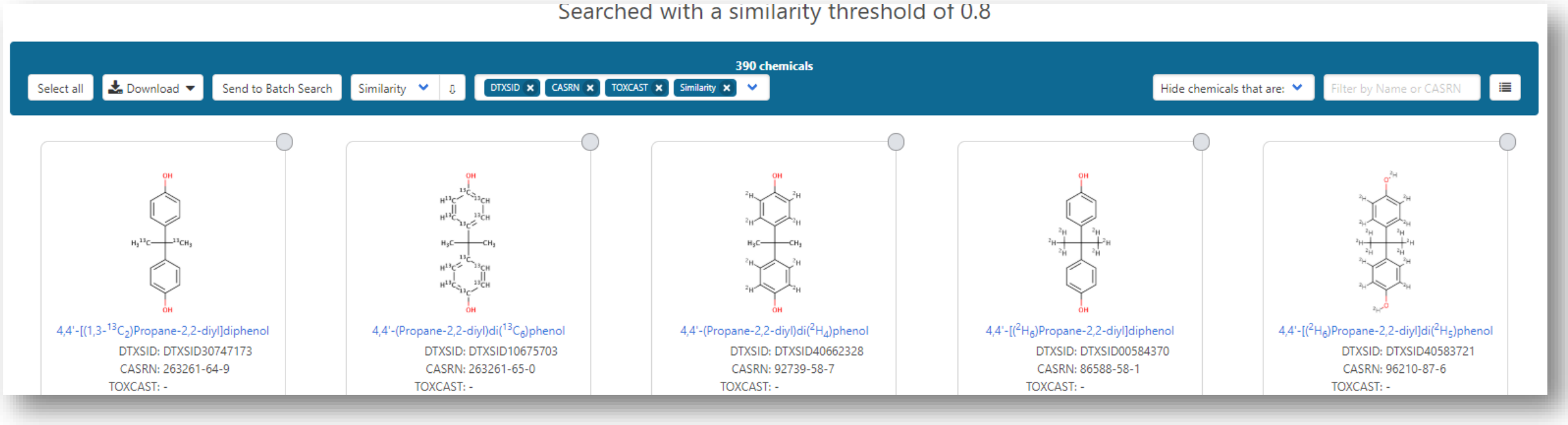
A new “Structure Zoom”

- On-click hover all over the dashboard as well as structure thumbnail



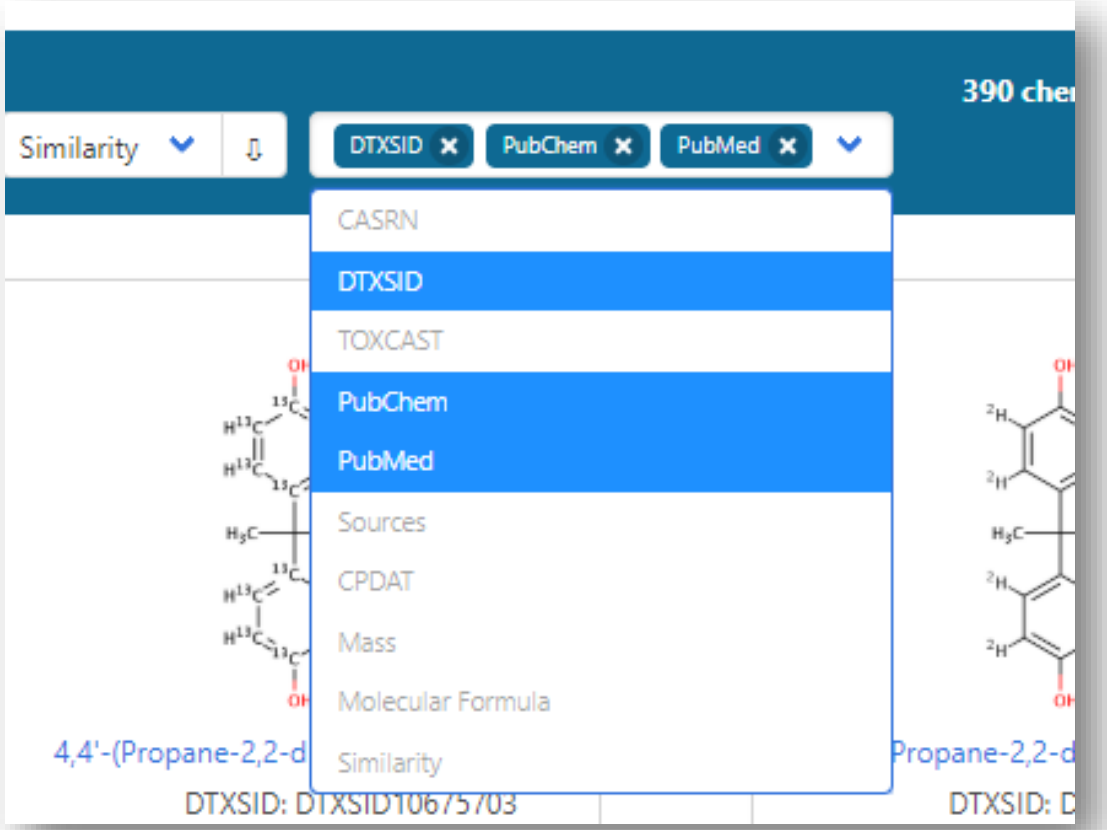


Reworked multiple results page



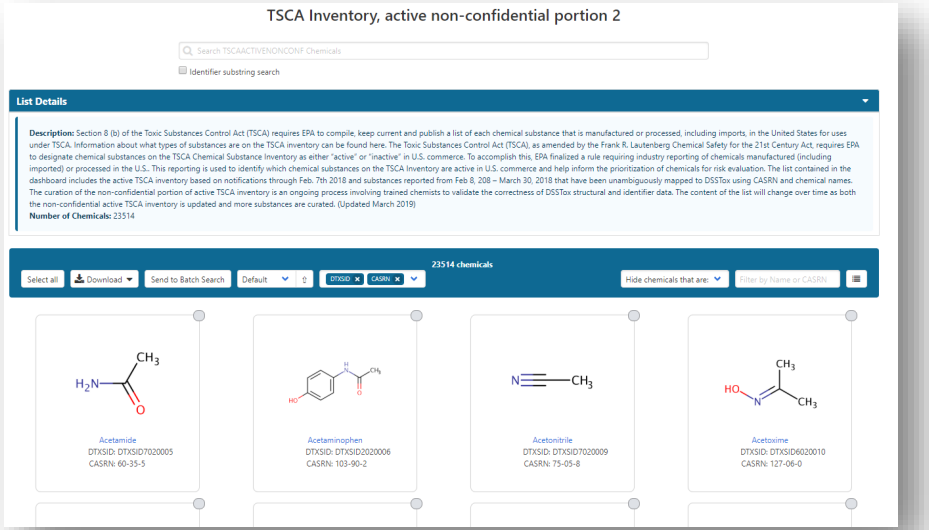
Reworked multiple results page

- Use Ctrl to select multiple display



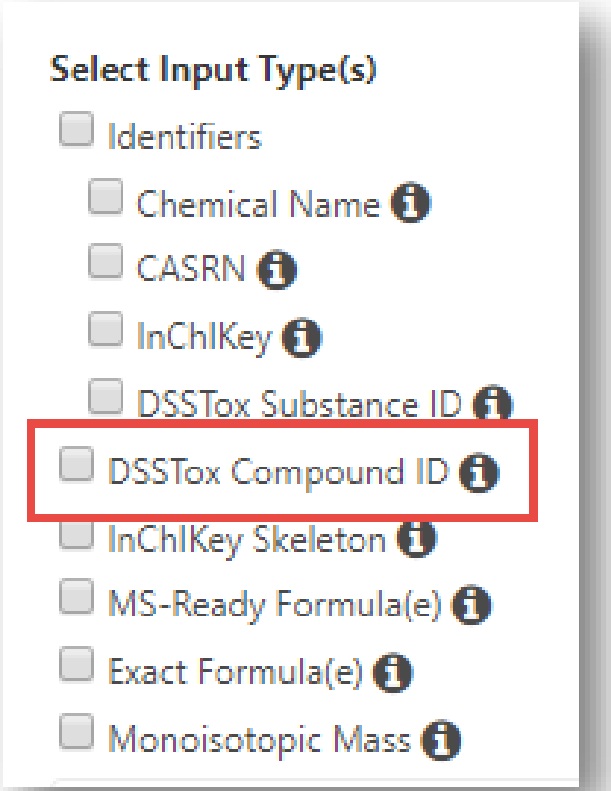
- Improved visual cues for loading large lists of chemicals

- Loading of Large lists RETAINS ordering

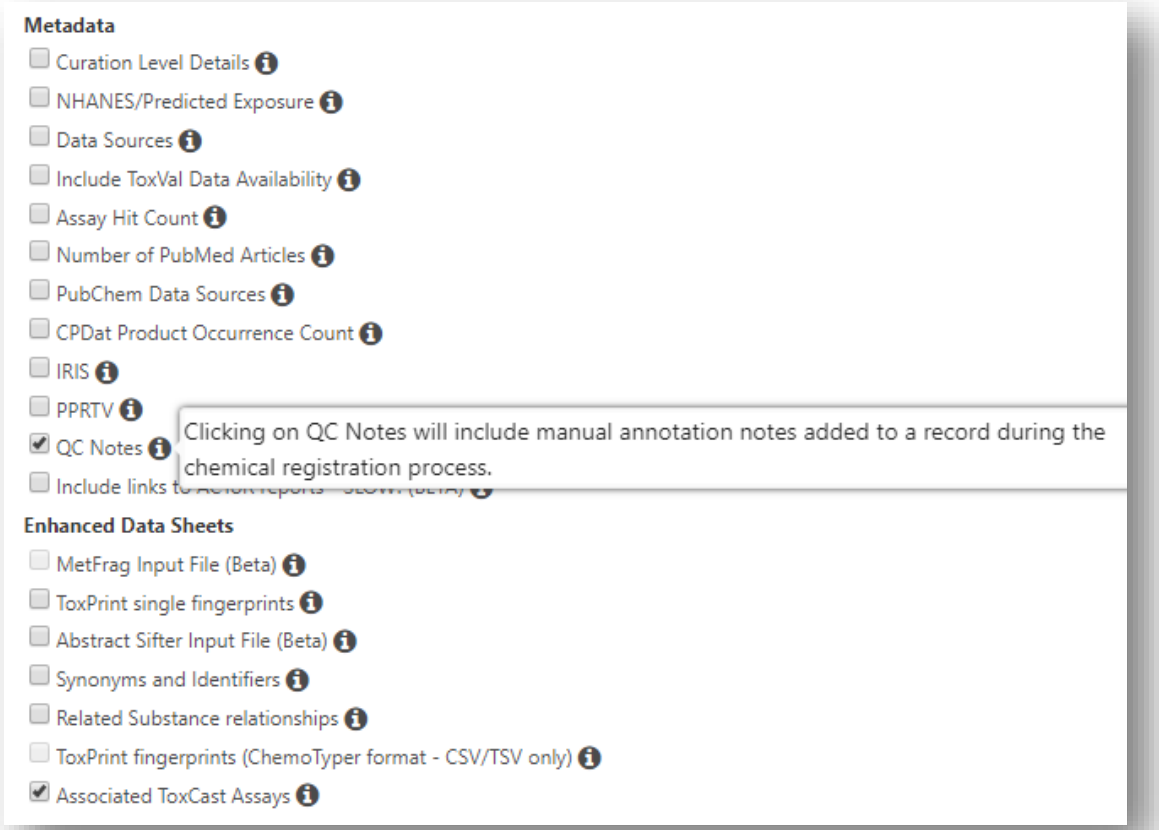


Batch Search

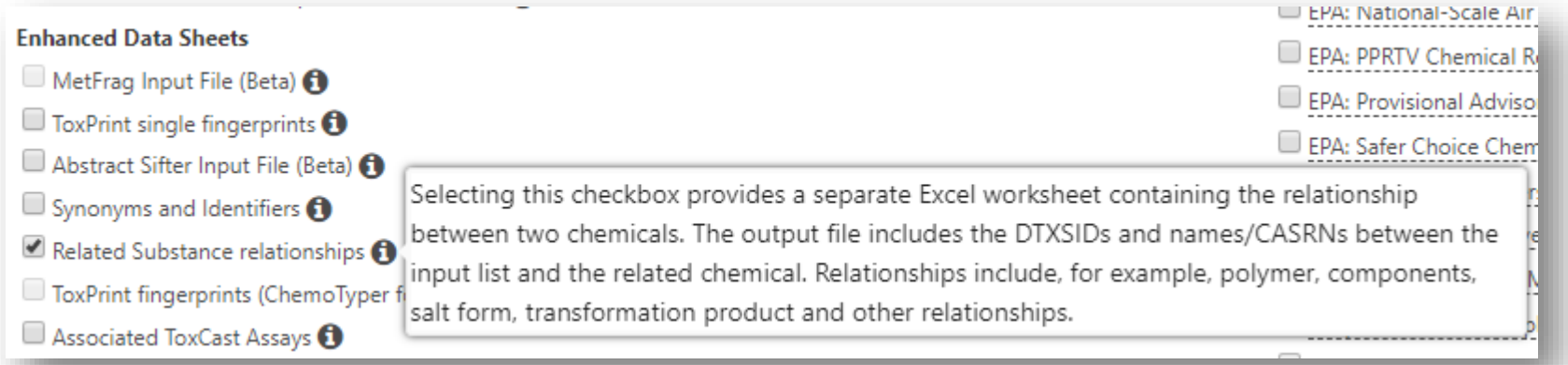
- New Search input - DTXCID



- New Search Outputs

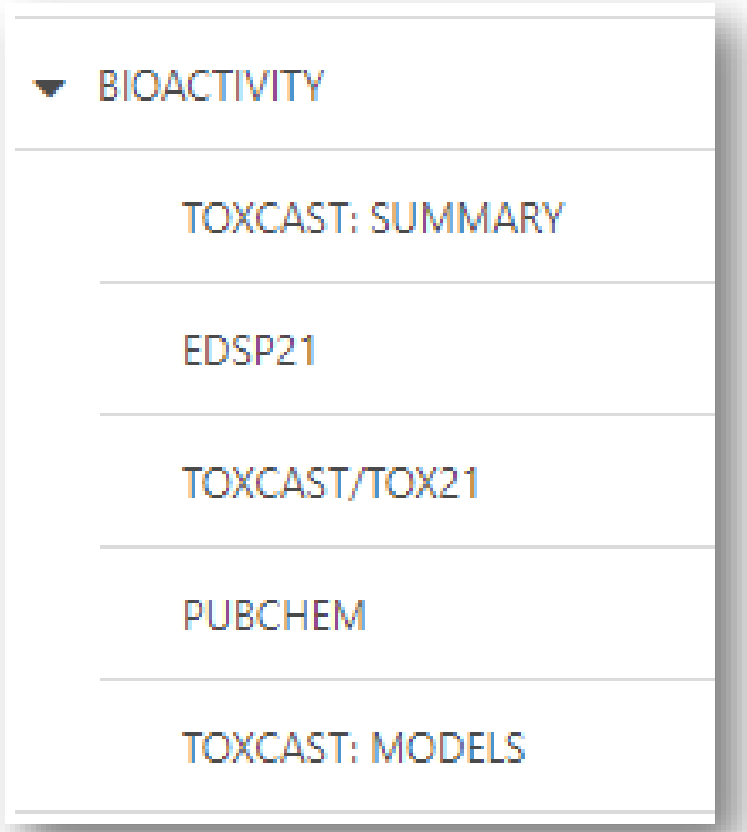


Related Substance Relationships



	A	B	C	D	E	F	G	H
1	INPUT	DTXSID	PREFERRED_NAME	HAS_RELATIONSHIP_WITH	RELATED_DTXSID	RELATED_PREFERRED_NAME	RELATED_CASRN	
2	xylenes	DTXSID2021446	Xylenes	Transformation Product	DTXSID40176394	N-Benzoylalanine	2198-64-3	
3	xylenes	DTXSID2021446	Xylenes	Component	DTXSID6026298	m-Xylene	108-38-3	
4	xylenes	DTXSID2021446	Xylenes	Component	DTXSID3021807	o-Xylene	95-47-6	
5	xylenes	DTXSID2021446	Xylenes	Component	DTXSID2021868	p-Xylene	106-42-3	
6	xylenes	DTXSID2021446	Xylenes	Predecessor: Component	DTXSID9021421	Xylenes; defined mixture 1	NOCAS_21421	
7	xylenes	DTXSID2021446	Xylenes	Predecessor: Component	DTXSID7021447	Xylenes; defined mixture 2	NOCAS_21447	
8	xylenes	DTXSID2021446	Xylenes	Predecessor: Component	DTXSID30891529	Total Petroleum Hydrocarbons (TPH)	NOCAS_891529	
9	xylenes	DTXSID2021446	Xylenes	Markush Child	DTXSID3021807	o-Xylene	95-47-6	
10	xylenes	DTXSID2021446	Xylenes	Markush Child	DTXSID6026298	m-Xylene	108-38-3	
11	xylenes	DTXSID2021446	Xylenes	Markush Child	DTXSID2021868	p-Xylene	106-42-3	

Bioactivity Data

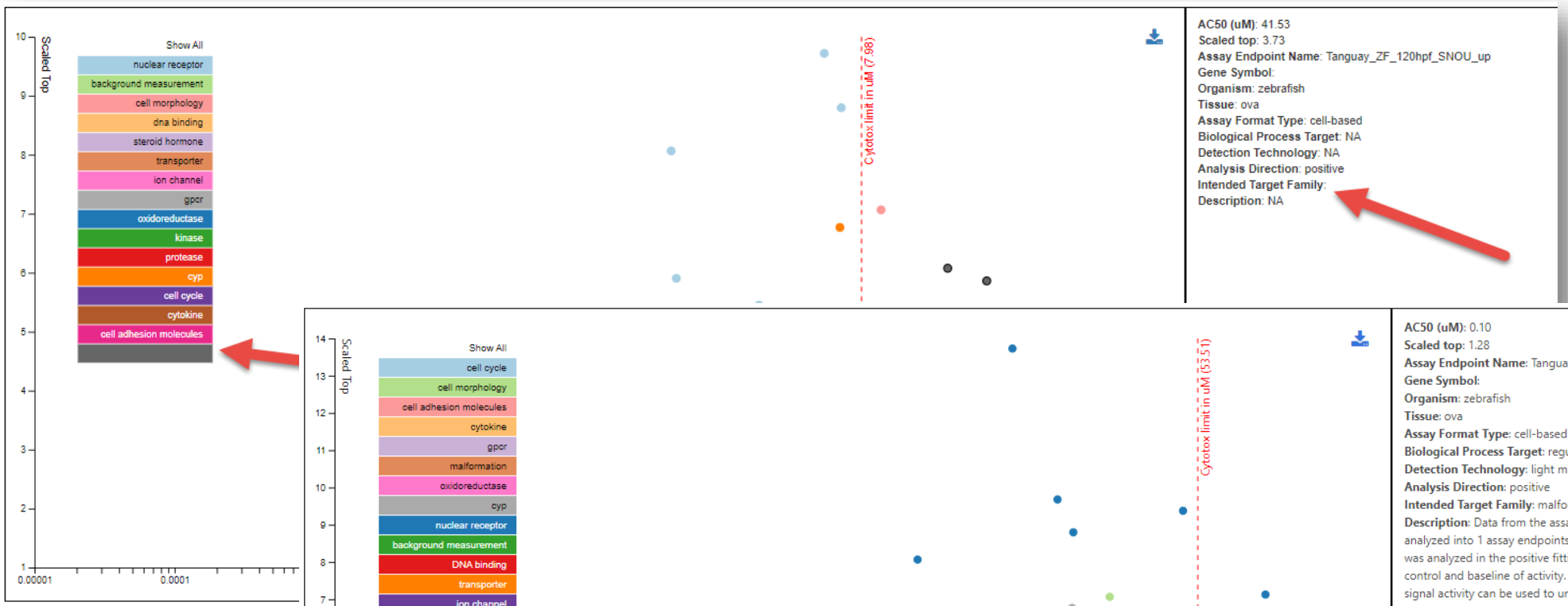


- Summary data now has “enhanced tables”
- EDSP21 subset of assays has grown
- Toxcast/Tox21 “all data” has been integrated
- PubChem data widget – no change
- Subset of ToxCast ”Models” – extended to include “COMPARA” data

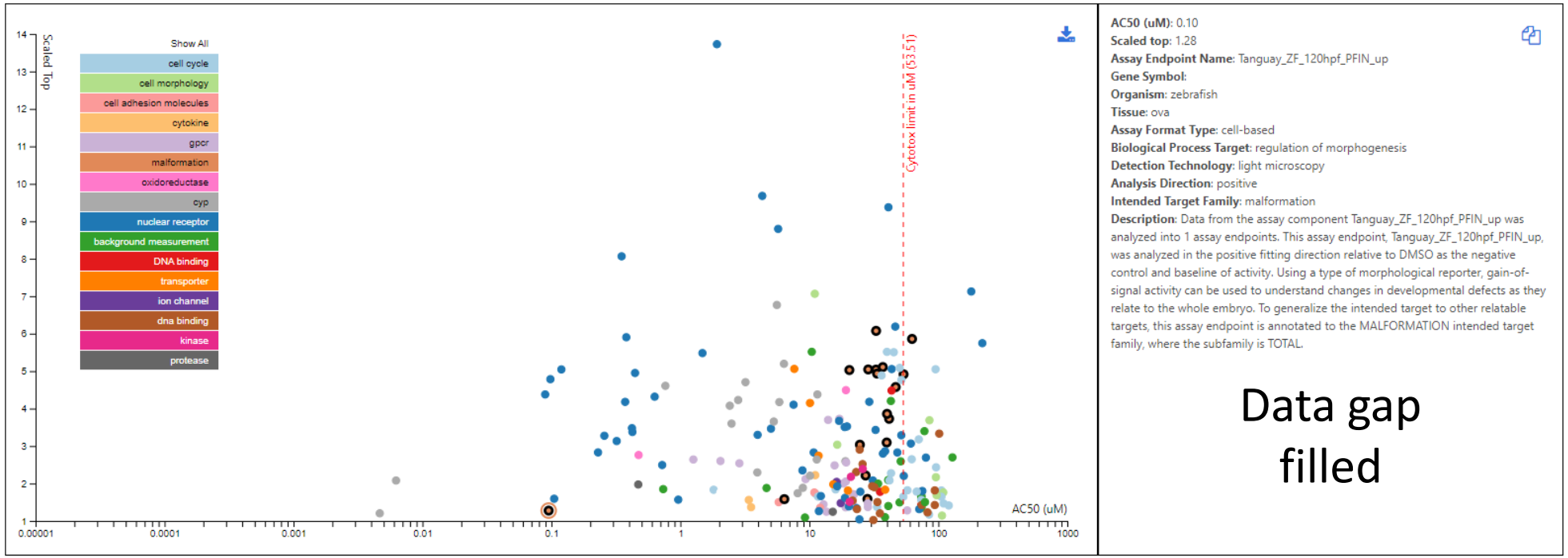
Tables Reworked – Column Selection

- Ability to select columns to show added for tables –Bioactivity most important – Pick your own preferred display

Assay Annotations Filled a Lot of Gaps!

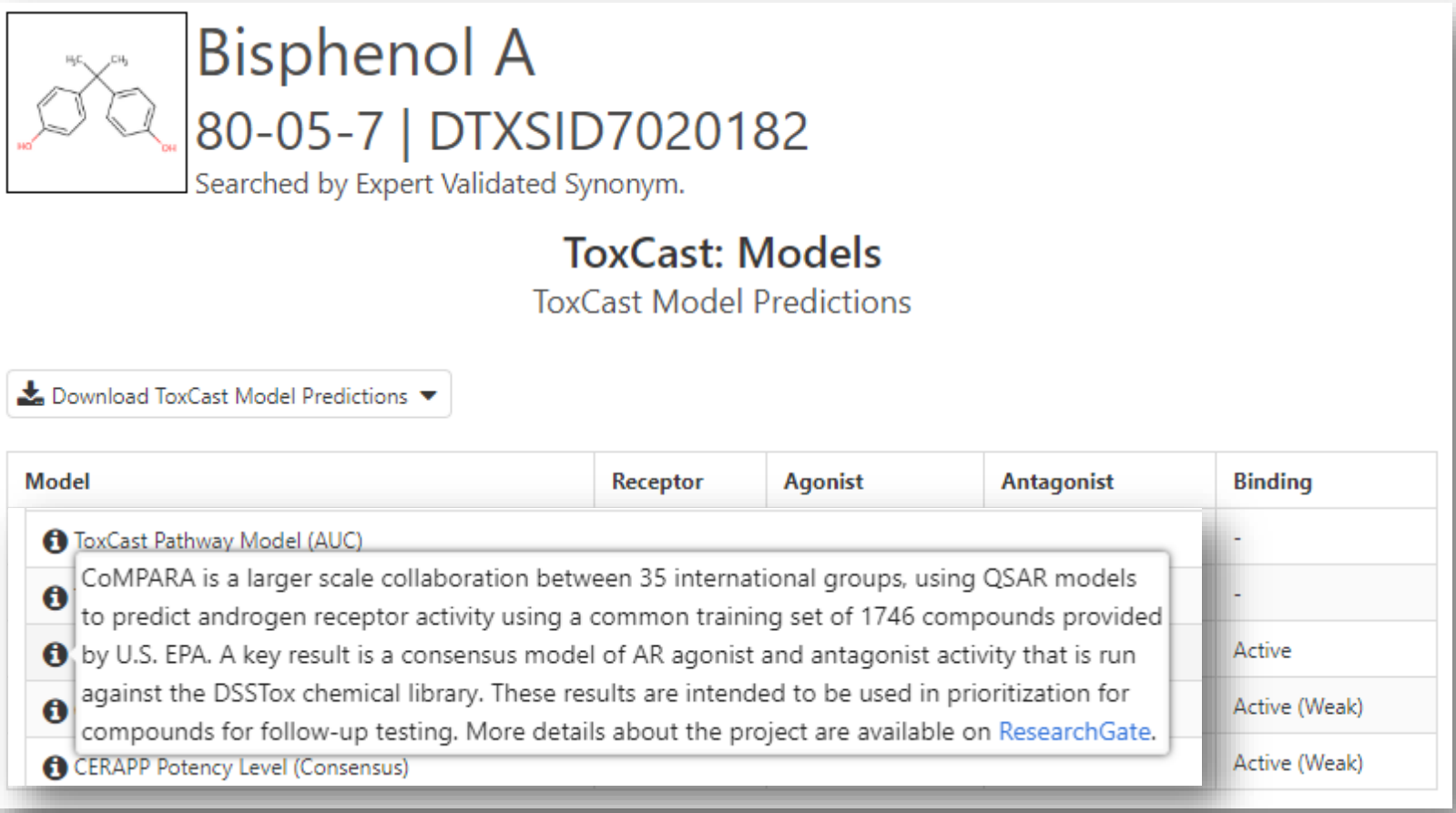


Previous data gaps



Data gap filled

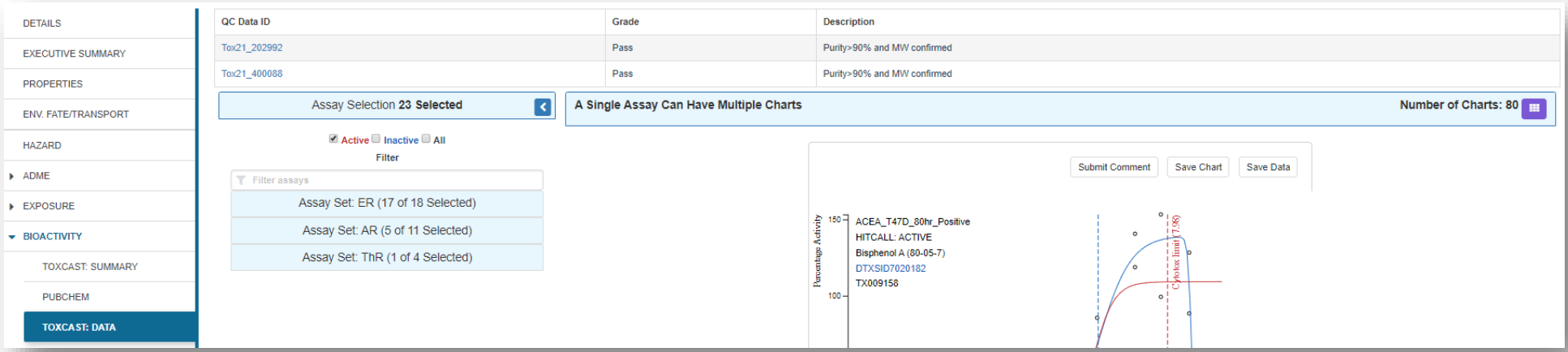
Toxcast: Models – COMPARA added



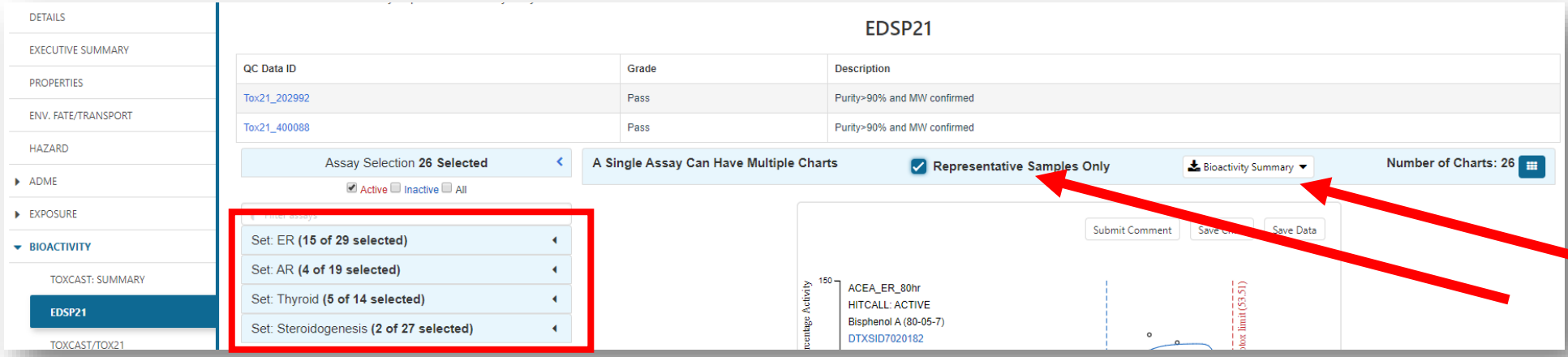
“EDSP Subset”

- New assays added – expanded all subsets. New set of steroidogenesis assays – including CEETOX data

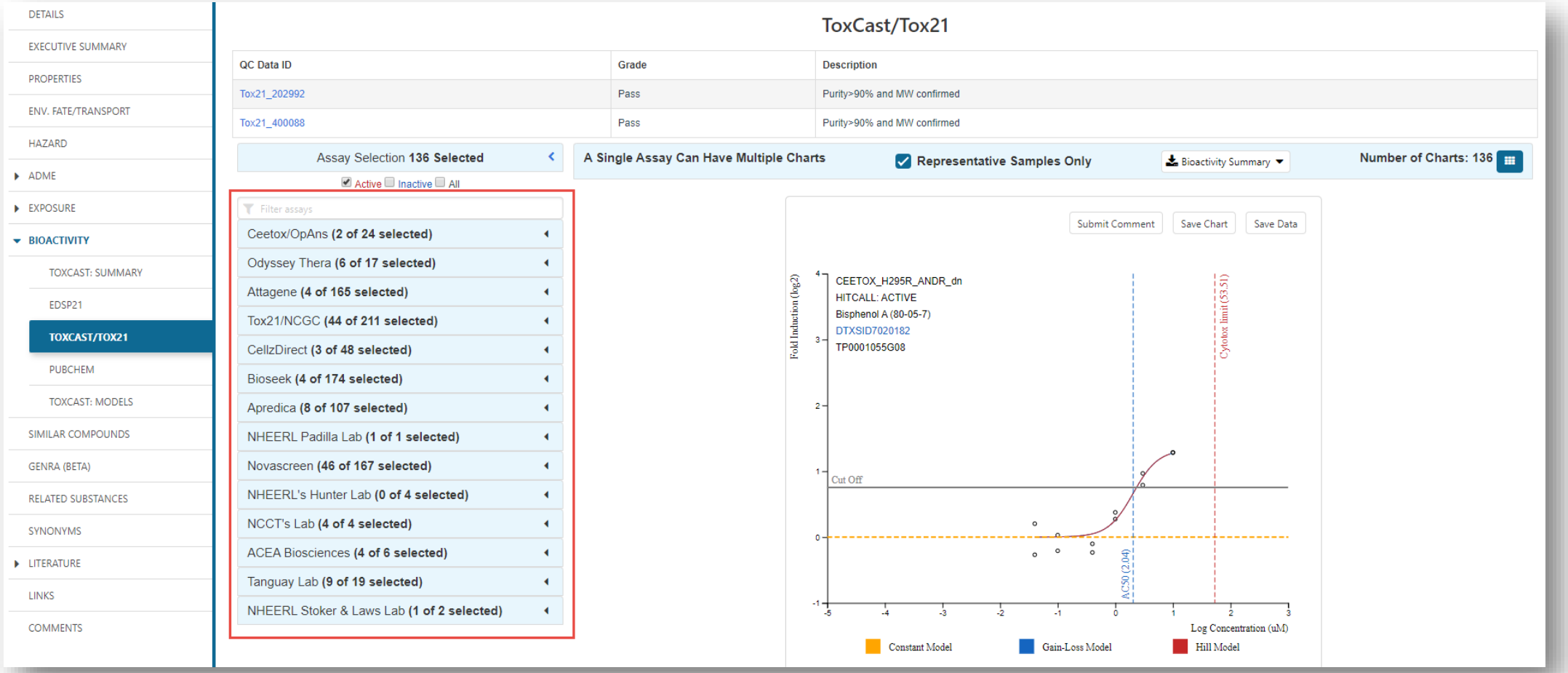
- Previous



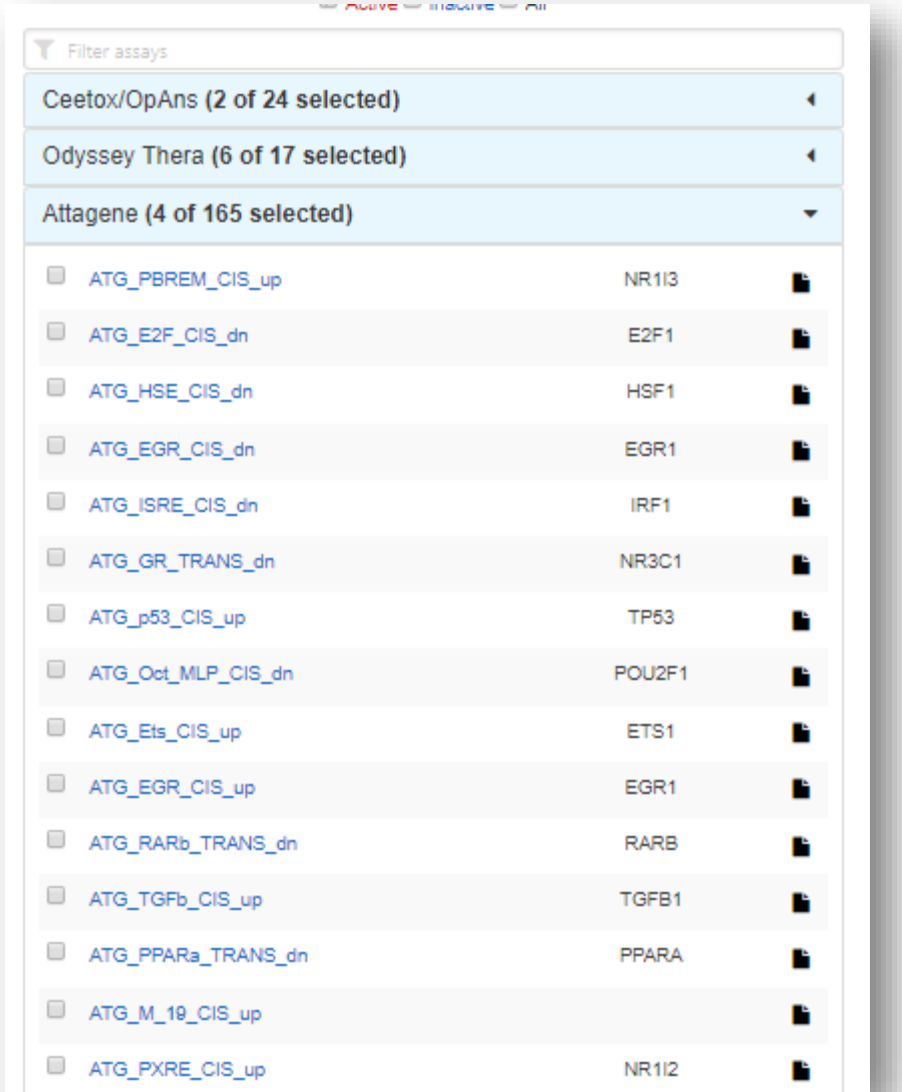
- Update

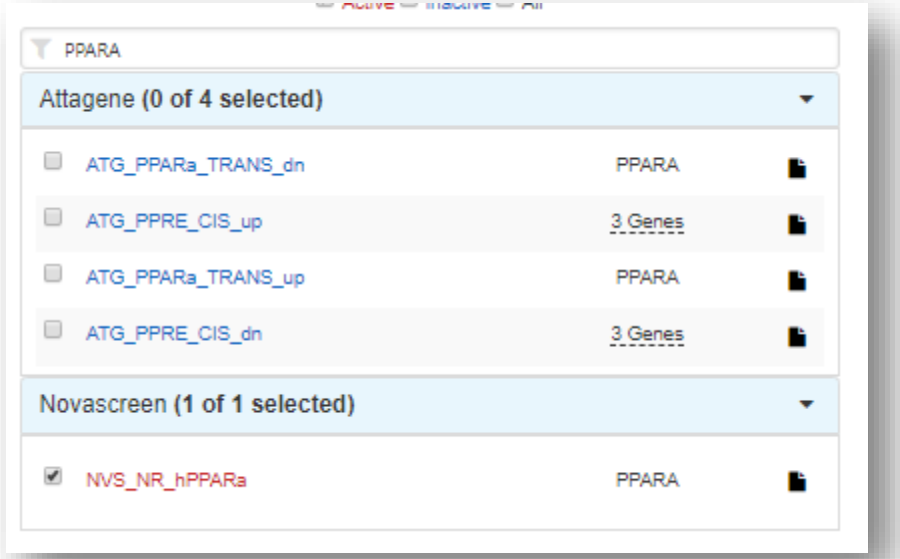


ToxCast/Tox21 Data – All data from invitroDBv3

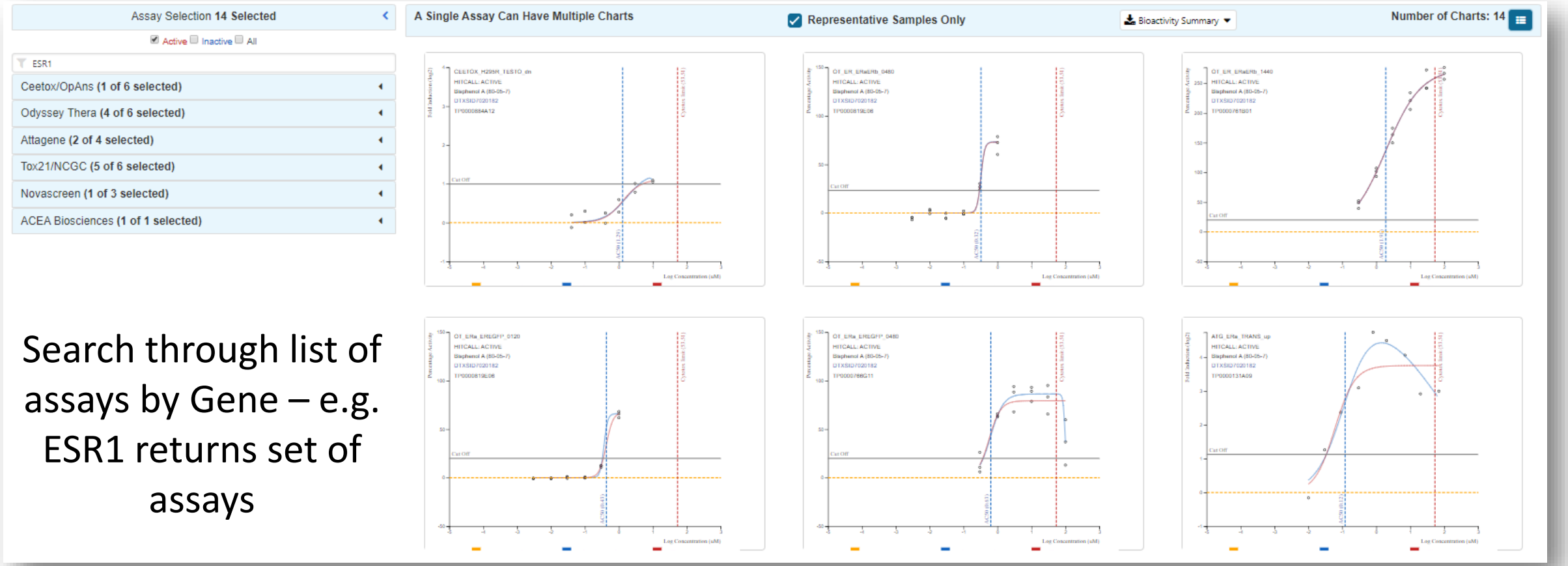


Filtering – Gene annotation added





Filtering



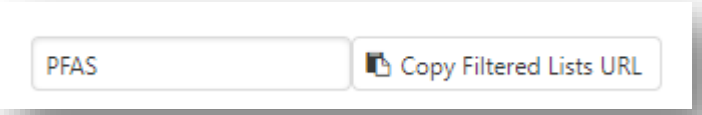
Search through list of assays by Gene – e.g. ESR1 returns set of assays

Lists of Chemicals/Lists of Assays

- Reworked Chemical List page – lots of lists added including segregation
  - LIST: Algal Toxins, Amino Acids, Bisphenol Compounds, PAHs, Synthetic Cannabinoids and Psychoactives, Vitamins, PCBs, PBDEs, Hair Dyes
  - WIKILIST: Additives in Cigarettes, Extremely Hazardous Substances
  - EPA: PALs, HPV list, Chemical Contaminants, PPRTV Reports etc, Pesticides Chemical Search
- Helps cluster in the batch search and as a query on the lists page (see figure)
- invitroDbv3 assays added to assay list

Lists of Chemicals

- Download the “list of lists” as Excel or TSV
- Subset of lists from query – “what are all PFAS lists?”



http://comptox.epa.gov/dashboard/chemical_lists/?search=PFAS

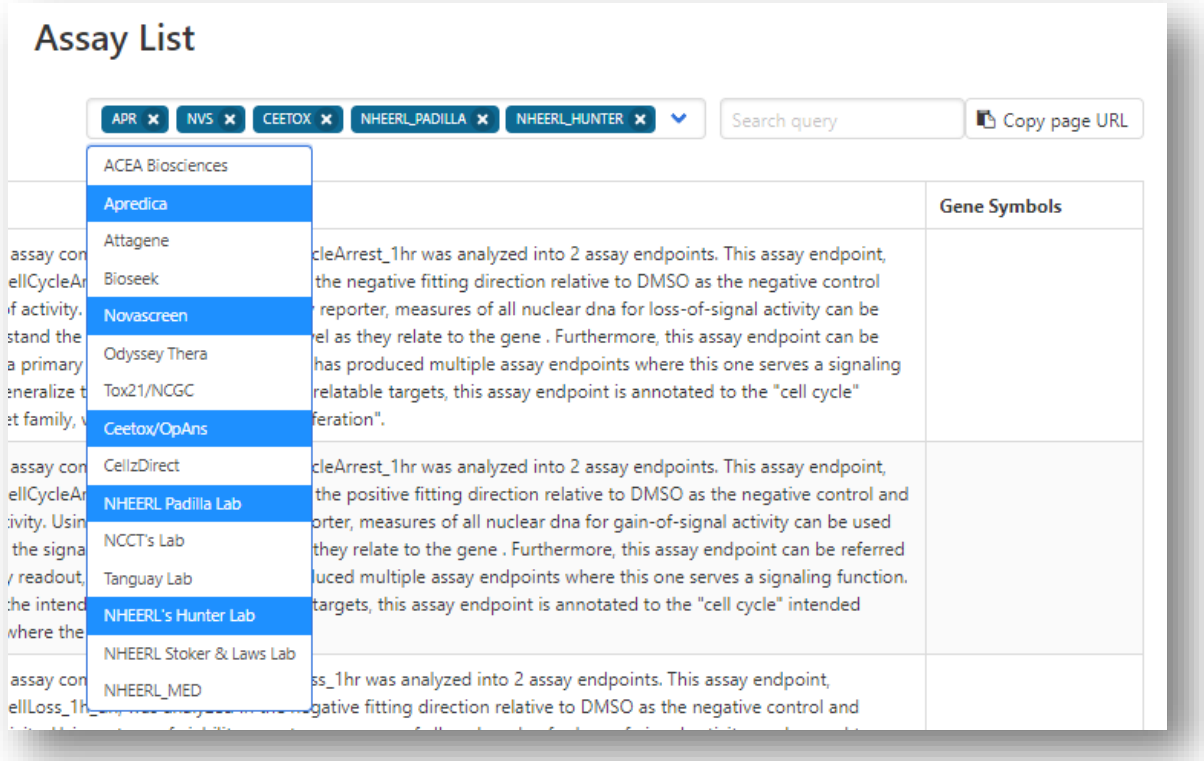
Select List

Download | Columns | 10 | Search query | Copy page URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
40CFR355	40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities	2018-01-05	354	Extremely Hazardous Substance List and Threshold Planning Quantities; Emergency Planning and Release Notification Requirements; Final Rule. (52 FR 13378)
ACSREAG	LIST: ACS Reagent Chemicals	2017-04-14	405	The ACS Committee on Analytical Reagents sets purity specifications for almost 500 reagent chemicals and over 500 standard-grade reference materials.
AEGLVALUES	AEGLS: Acute Exposure Guideline Levels	2018-04-20	174	Acute exposure guideline levels (AEGLs) describe the human health effects from once-in-a-lifetime, or rare, exposure to airborne chemicals.
ALGALTOX	LIST: Algal Toxins	2018-05-04	54	A list of Algal Toxins of potential interest
AMINOACIDS	LIST: Amino Acids	2019-02-04	0	Amino acids are organic compounds containing amine (-NH2) and carboxyl (-COOH) functional groups, along with a side chain (R group) specific to each amino acid.
APCRA_PRO	LIST: APCRA Chemicals for Prospective Analysis	2018-02-14	204	The APCRA prospective case study list of approximately 200 chemicals as of January 2018, developed by ECHA in consultation with EPA and other partners

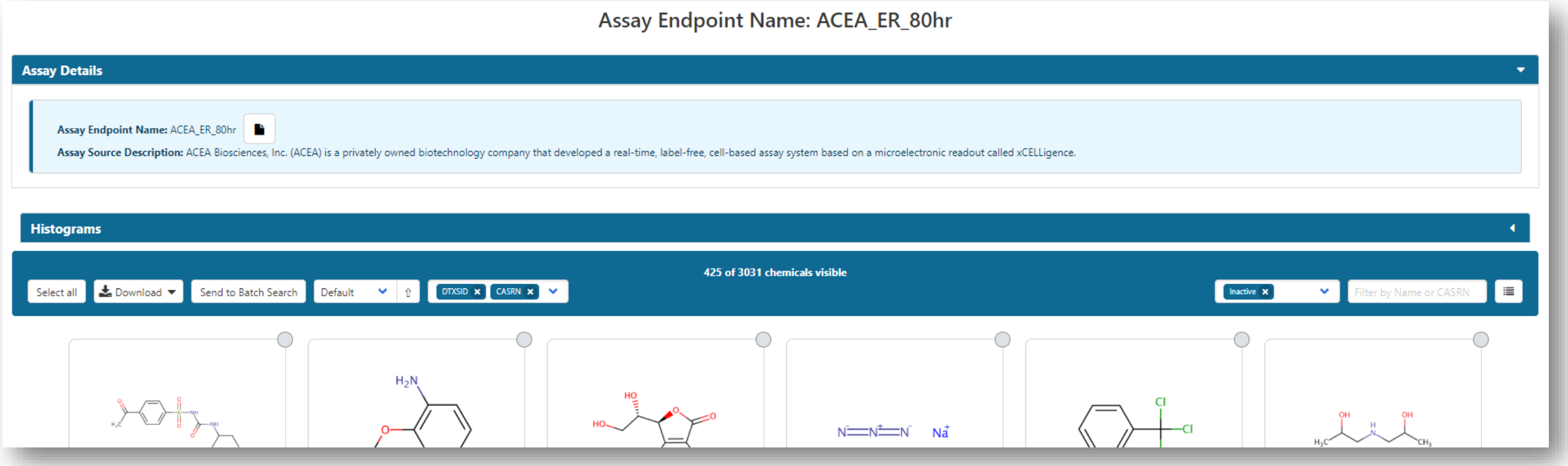
List of Assays

- Download list of all assays (Excel or TSV)
- Filter by Vendor or Multiple Vendors
- Subset of lists from query based on substring search e.g. http://comptox.epa.gov/dashboard/assay_endpoints/?search=ESR1



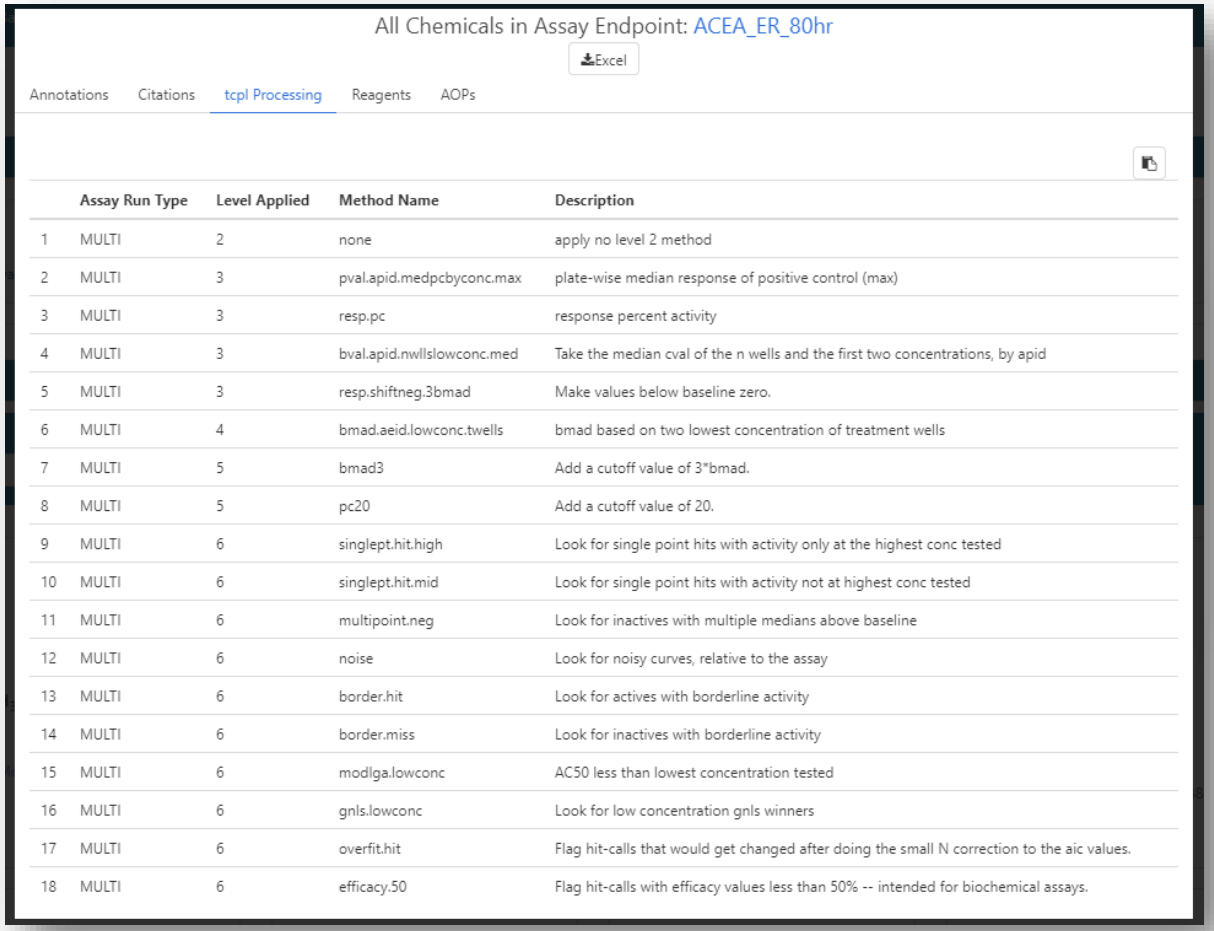
For a Specific Assay List

- All the advantages of the new “Multiple Results Page” plus…



Specific Assay List

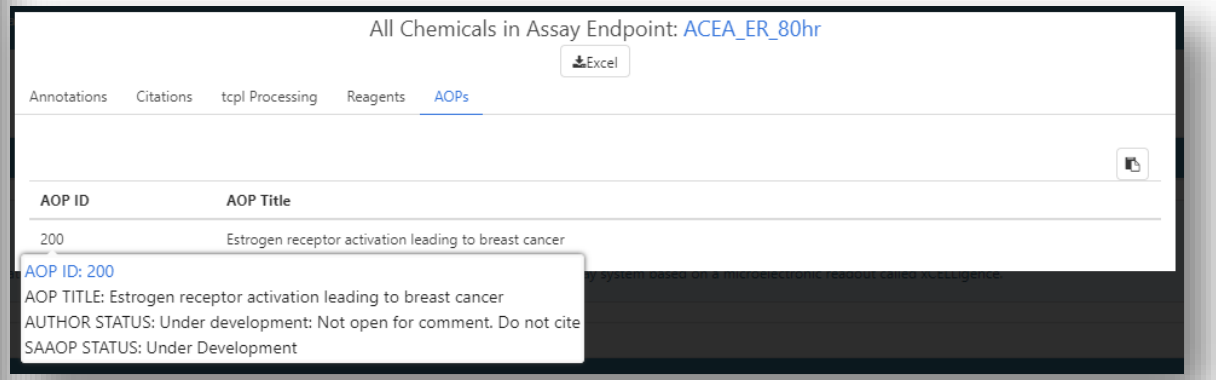
- Reworked assay table – more details available including AOP Wiki link


All Chemicals in Assay Endpoint: ACEA_ER_80hr

Excel

Annotations | Citations | tcpl Processing | Reagents | AOPs

	Assay Run Type	Level Applied	Method Name	Description
1	MULTI	2	none	apply no level 2 method
2	MULTI	3	pval.apid.medpcbyconc.max	plate-wise median response of positive control (max)
3	MULTI	3	resp.pc	response percent activity
4	MULTI	3	bval.apid.nwllslowconc.med	Take the median cval of the n wells and the first two concentrations, by apid
5	MULTI	3	resp.shiftneg.3bmad	Make values below baseline zero.
6	MULTI	4	bmad.aeid.lowconc.twells	bmad based on two lowest concentration of treatment wells
7	MULTI	5	bmad3	Add a cutoff value of 3*bmad.
8	MULTI	5	pc20	Add a cutoff value of 20.
9	MULTI	6	singlept.hit.high	Look for single point hits with activity only at the highest conc tested
10	MULTI	6	singlept.hit.mid	Look for single point hits with activity not at highest conc tested
11	MULTI	6	multipoint.neg	Look for inactives with multiple medians above baseline
12	MULTI	6	noise	Look for noisy curves, relative to the assay
13	MULTI	6	border.hit	Look for actives with borderline activity
14	MULTI	6	border.miss	Look for inactives with borderline activity
15	MULTI	6	modlga.lowconc	AC50 less than lowest concentration tested
16	MULTI	6	gnls.lowconc	Look for low concentration gnls winners
17	MULTI	6	overfit.hit	Flag hit-calls that would get changed after doing the small N correction to the aic values.
18	MULTI	6	efficacy.50	Flag hit-calls with efficacy values less than 50% -- intended for biochemical assays.


All Chemicals in Assay Endpoint: ACEA_ER_80hr

Excel

Annotations | Citations | tcpl Processing | Reagents | AOPs

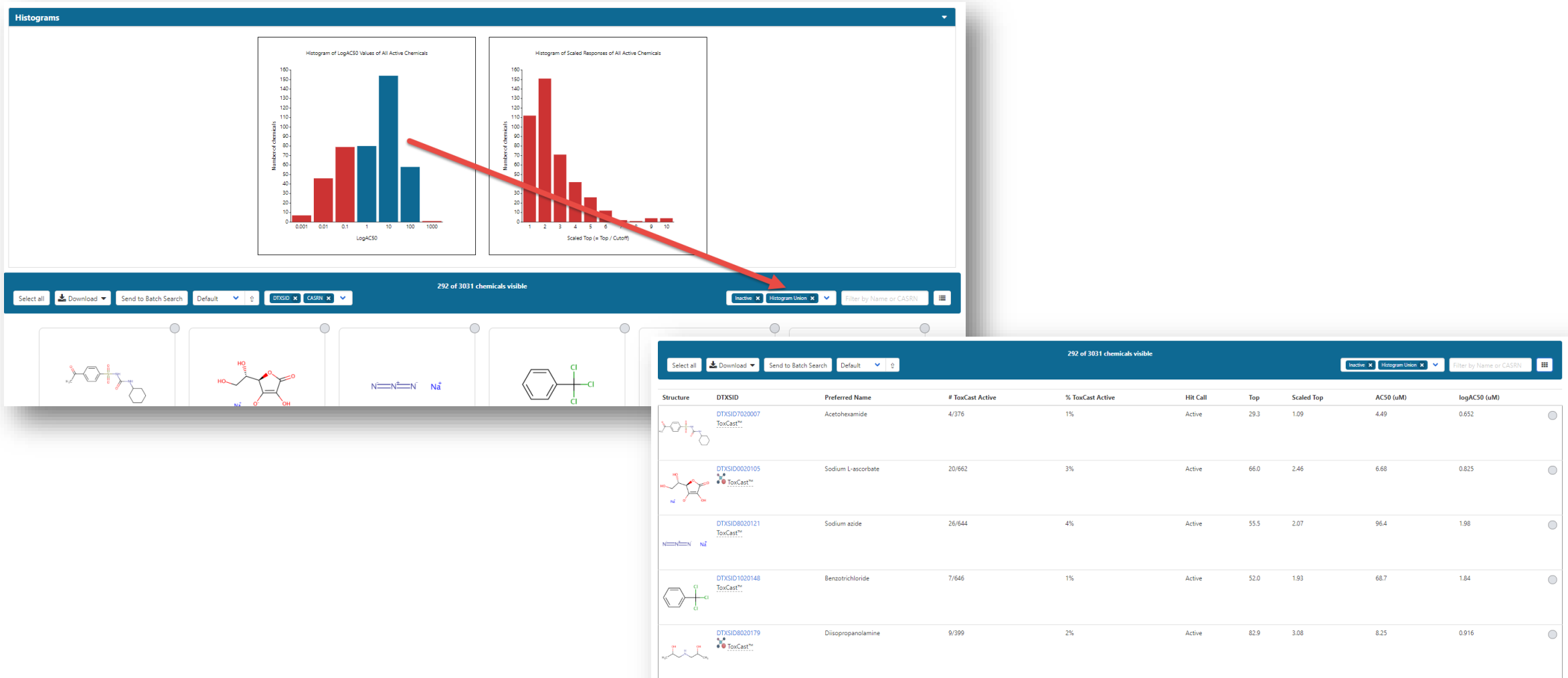
AOP ID	AOP Title
200	Estrogen receptor activation leading to breast cancer

AOP ID: 200
AOP TITLE: Estrogen receptor activation leading to breast cancer
AUTHOR STATUS: Under development: Not open for comment. Do not cite
SAAOP STATUS: Under Development

Specific Assay List – Histogram summary view

Specific Assay List – Histogram summary view

- Display specific subset of data from histogram – switch to Table Mode



The Underlying ToxCast data: Release of invitrodb, version 3.1

Accessing data downloads via FTP

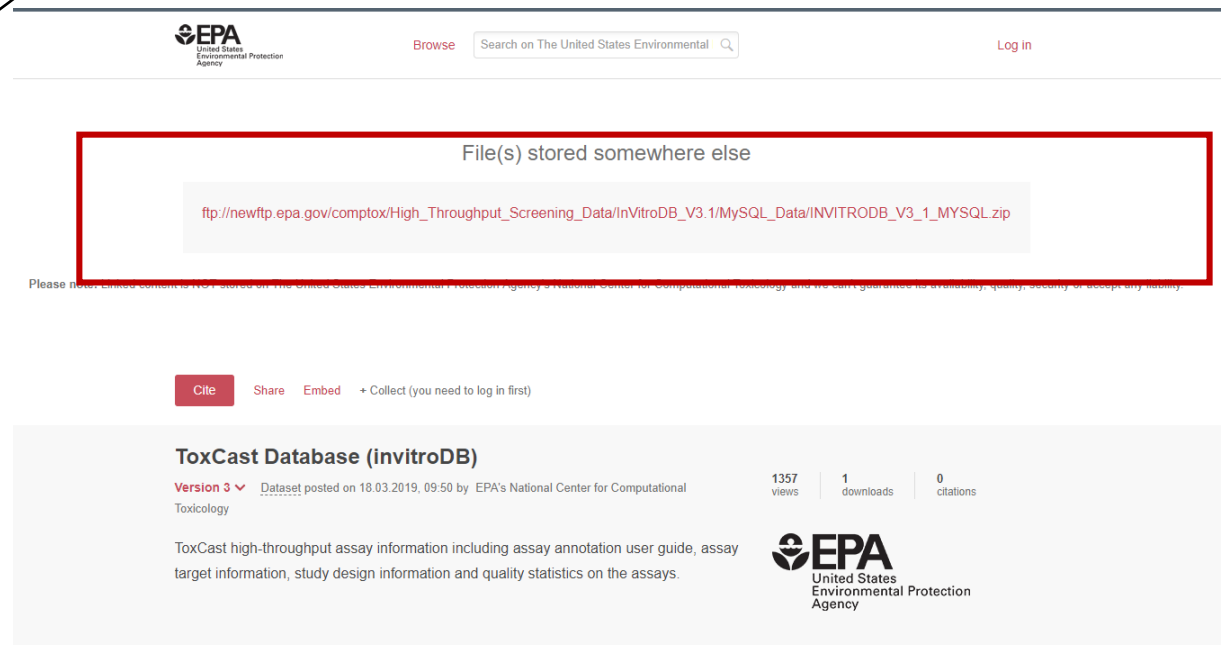
https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data

Click here for SQL of invitrodb_v3.1

ToxCast Data and Information

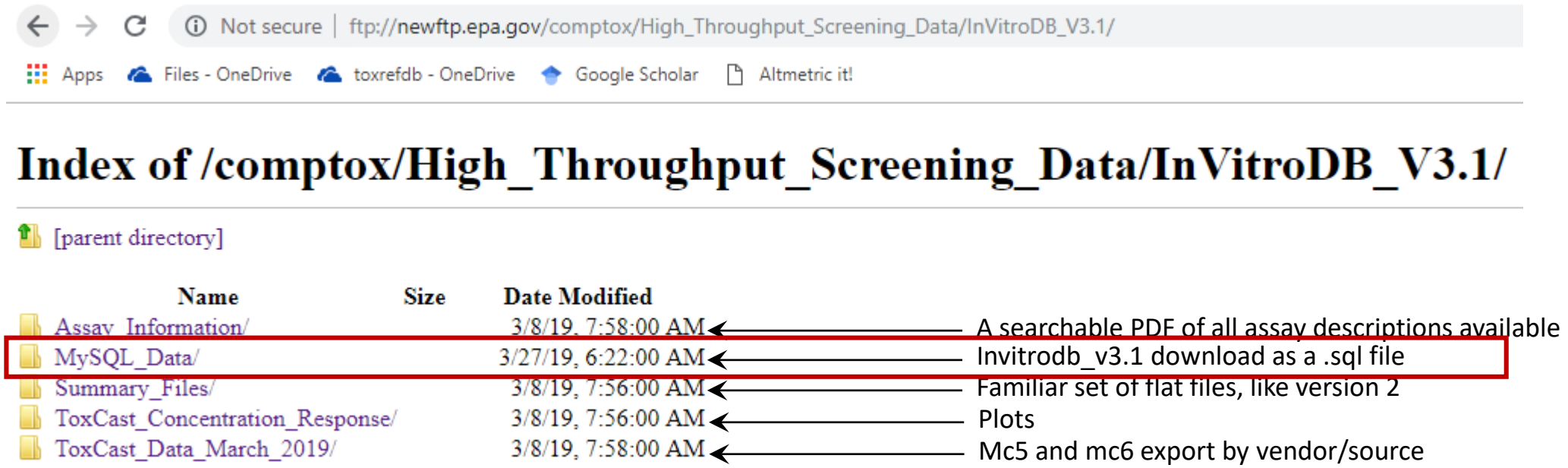
- About ToxCast

- **ToxCast & Tox21 Summary Files for invitroDBv3.1**. Data for a single chemical endpoint pair for thousands of chemicals and assay endpoints for 20 variables such as the activity or hit call, activity concentrations, whether the chemical was tested in a specific assay, etc.
  - Download ToxCast Summary Information
  - Download ReadMe
- **ToxCast & Tox21 Data Spreadsheet for invitroDBv3.1**. A spreadsheet of EPA's analysis of the chemicals screened through ToxCast and the Tox21 collaboration which includes EPA's activity calls from the screening of over 1,800 chemicals.
  - Download Data
  - Download ReadMe
- **ToxCast Data Pipeline R Package (tcpl released with invitroDBv3.1).** The R computer programming package used to process and model all EPA ToxCast and Tox21 chemical screening data. The files include the R programming package as well as documents that provide overviews of the data analysis pipeline used and the R package. Users will need experience with R to use these files.
  - Download Package
  - TCPL Overview
- **ToxCast Database (invitroDBv3.1)**: Database of EPA's analysis of chemicals screened through ToxCast assays, includes a MySQL database and the R package used to process the data.
  - Download Database for MAC
  - Download Database for Windows
  - Download ReadMe
- **ToxCast & Tox21 Concentration Response Plots for invitroDBv3.1**. Concentration response plots for all of the ToxCast and Tox21 assays.
  - Download Concentration Response Plots
- **ToxCast & Tox21 Chemicals Distributed Structure-Searchable Toxicity Database (DSSTox files) for invitroDBv3.** Chemical lists and information for 9,403 unique substances (DTXSIDs) and DSSTox standard chemical fields (chemical name; CASRN; structure; etc.). These chemicals have been evaluated through the ToxCast and Tox21 high-throughput screening efforts.
  - ToxCast Chemicals: Data Management and Quality Considerations Overview
  - Download ToxCast Chemical Information
  - Download ReadMe
- **ToxCast & Tox21 High-Throughput Assay Documentation.** ToxCast high-throughput assay documentation including descriptions, target information, study design information and quality statistics.
  - Assay Descriptions **(work in progress)** - Descriptions and guidelines for ToxCast assays in format outlined by the OECD Guidance Document 211 for describing non-guideline in vitro test methods. The intent of GD 211 is to harmonize non-guideline, in vitro method descriptions to allow assessment of the relevance of the test method for biological responses of interest and the quality of the data produced. This document is organized by assay platform providers. You can also find descriptions for endocrine-related assays. It is a



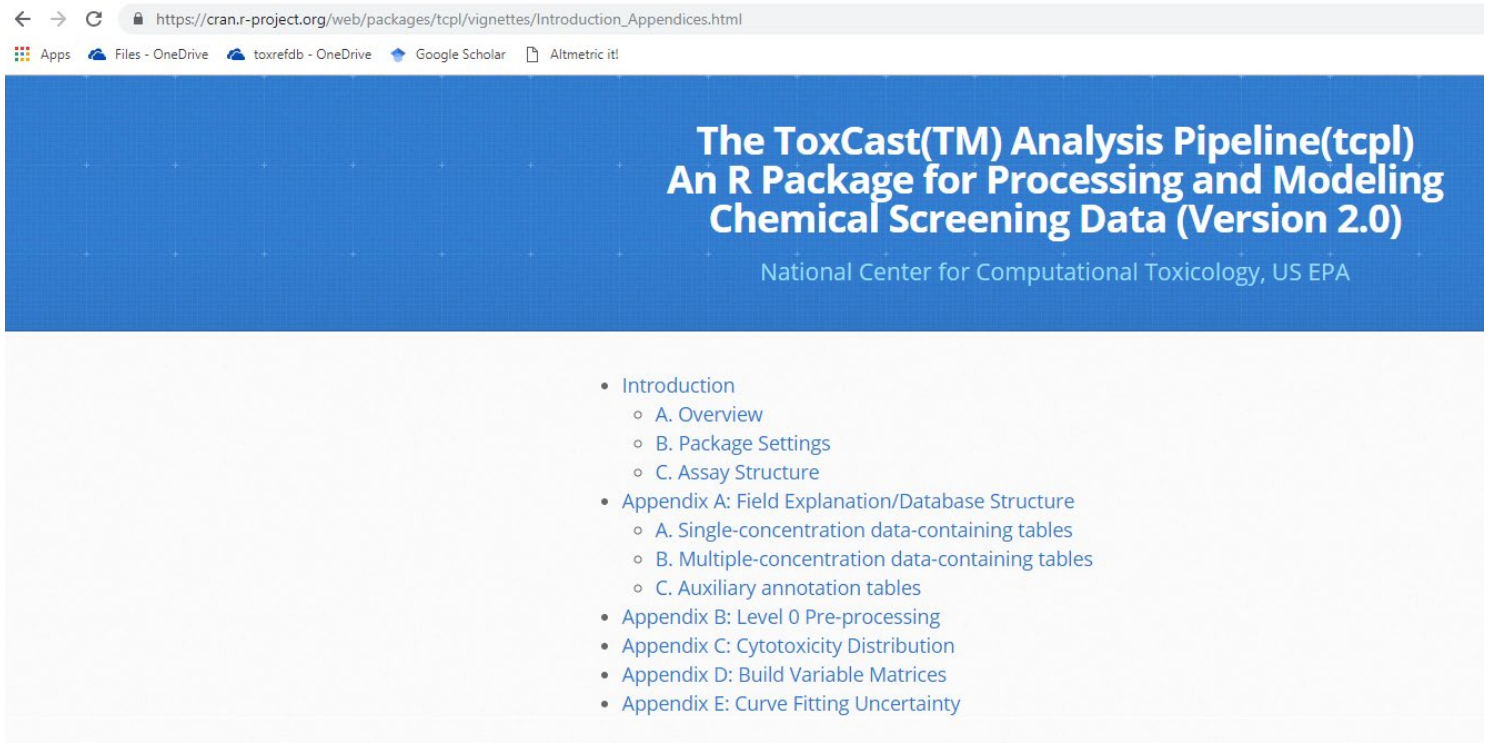
Current snapshots, March 28, 2019

Accessing invitrodb_v3.1 download via FTP



Extracting information from invitrodb_v3.1

- Can use tcpl version 2.0.1 live now on CRAN https://cran.r-project.org/web/packages/tcpl/
- Refer to our rewritten vignettes that explain invitrodb and the ToxCast Pipeline, example:



Additional tables in invitrodb_v3.1

- Cytotox includes the “burst” information
- EDSP21-related models: model_generic_chemical_ar_scores, model_generic_chemical_cerapp_score, model_generic_chemical_compara_scores, model_generic_chemical_er_scores, model_generic_chemical_hth295r_scores
- A number of tables aimed at increased assay annotation:
  - Structured assay description tables (assay_component_descriptions, assay_component_endpoint_descriptions, assay_descriptions)
  - Assay lists (assay_list and assay_list_aeid) used in research to group assays (e.g., endocrine-relevant assays for the EDSP21 part of the dashboard)
  - Assay ontology tables for future dashboard searching (ontology and ontology_invitrodb) based on Bioassay Ontology codes
  - Gene/intended target tables updated
- Mc7: uncertainty information obtained using *toxboot* (https://github.com/ericwatt/toxboot)

Steroidogenesis model information is further described in Haggard et al. 2018 (PMID 29216406).

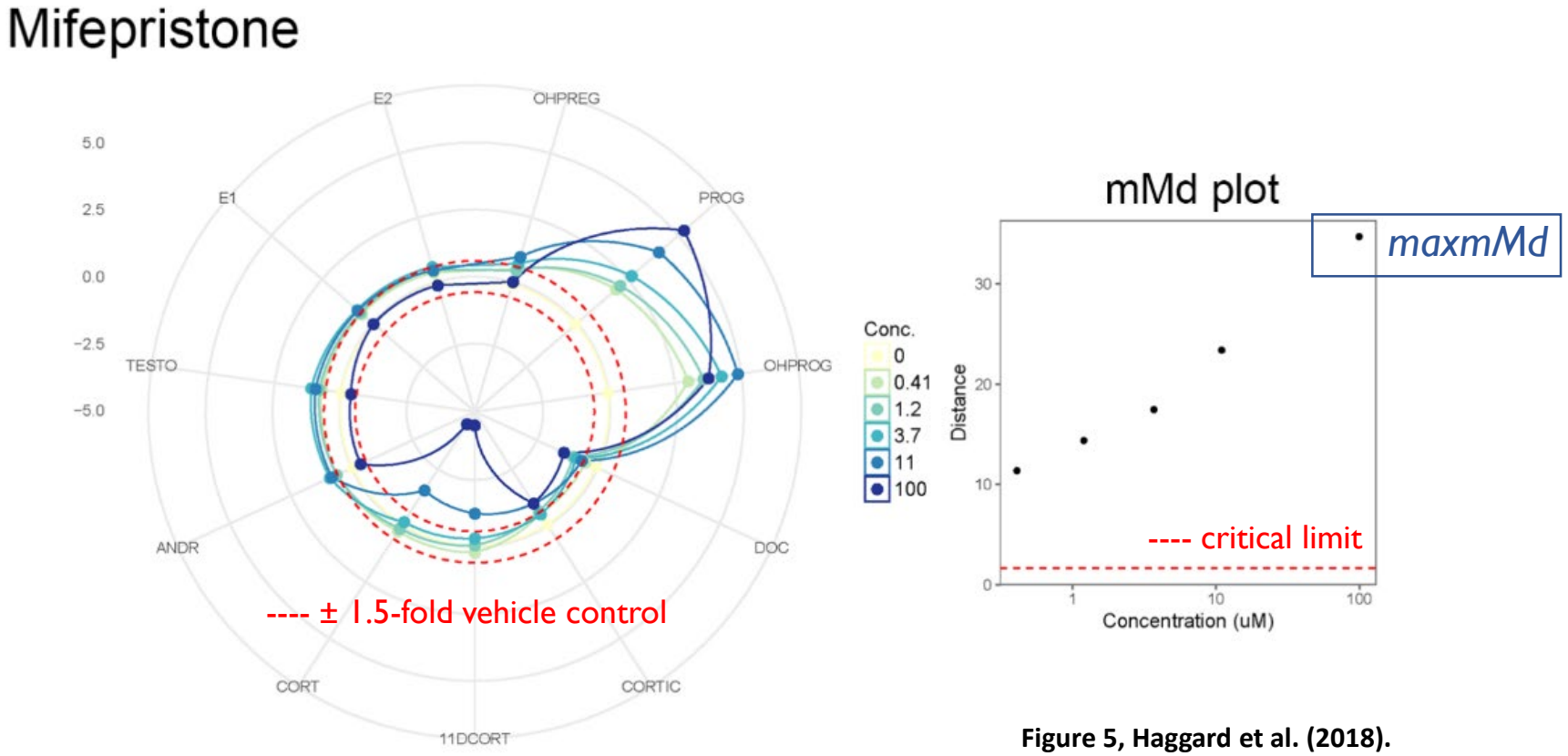


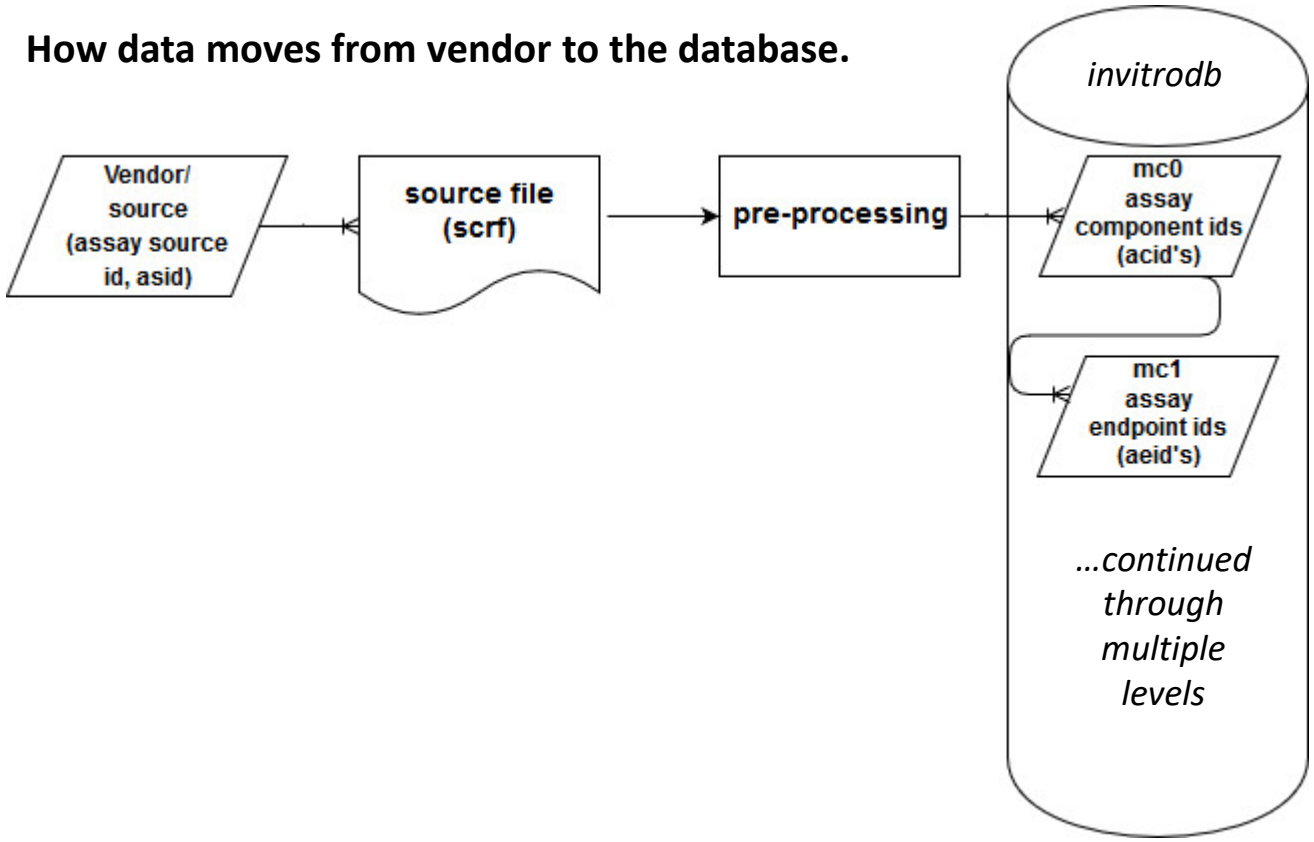
Figure 5, Haggard et al. (2018).

Mifepristone strongly modulated progestagens with significant effects on progesterone and OH-progesterone and moderate but non-significant trends on corticosteroids and androgens, resulting in a relatively high adjusted maxmMd of 33.

- Reduced an 11-dimensional question to a single dimension.
- Selection of the maxmMd appeared to provide a reproducible, quantitative approximation of the magnitude of effect on steroidogenesis.

Organization of data entering invitrodb

How data moves from vendor to the database.

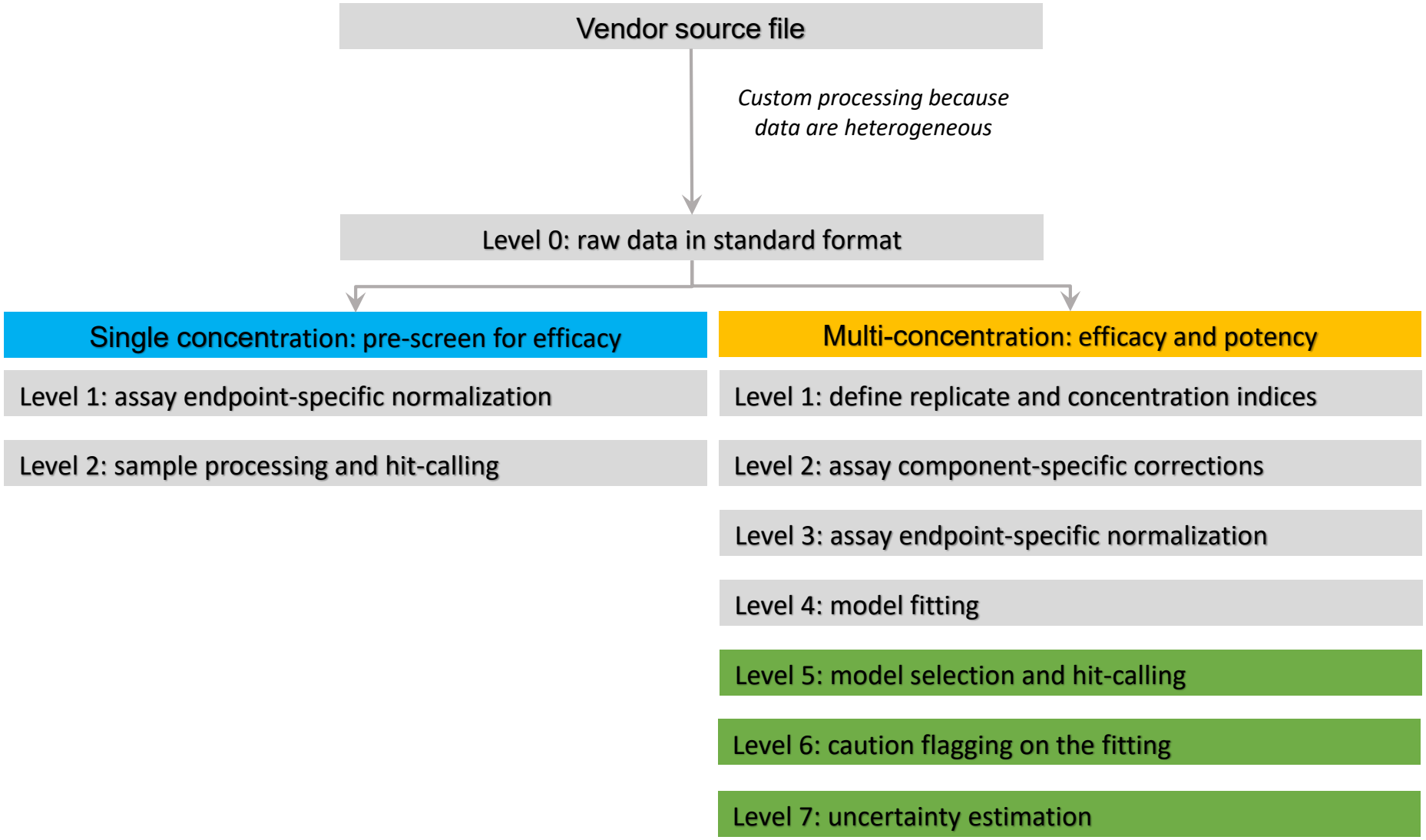


Example: asid to acid to aeid.
acid can be 1:1 or 1:many with aeid.

```
> tcplLoadAsid()
    asid           asnm
 1:    1           ACEA
 2:    2            APR
 3:    3            ATG
 4:    4            BSK
 5:    5            NVS
 6:    6             OT
 7:    7          TOX21
 8:    8         CEETOX
 9:   11            CLD
10:   12 NHEERL_PADILLA
11:   17   NCCT_SIMMONS
12:   13        TANGUAY
> tcplLoadAcid(fld='asid', val=8)
    asid acid                    acnm
 1:    8  586   CEETOX_H295R_11DCORT
 2:    8  587    CEETOX_H295R_OHPREG
 3:    8  588    CEETOX_H295R_OHPROG
 4:    8  589      CEETOX_H295R_ANDR
 5:    8  591  CEETOX_H295R_CORTISOL
 6:    8  593       CEETOX_H295R_DOC
 7:    8  594 CEETOX_H295R_ESTRADIOL
 8:    8  595   CEETOX_H295R_ESTRONE
 9:    8  597      CEETOX_H295R_PROG
10:    8  598     CEETOX_H295R_TESTO
> tcplLoadAeid(fld='acid', val=586)
   acid aeid                   aenm
1:  586  890 CEETOX_H295R_11DCORT_dn
2:  586  891 CEETOX_H295R_11DCORT_up
```

- Assay sources or vendors may send many files, which are pre-processed.
- The mc0 data in invitrodb is at the assay component level.
- At mc1, assay endpoints are defined, but it is not until normalization at mc3 that data are retrieved by assay endpoint.

Outline of the ToxCast pipeline

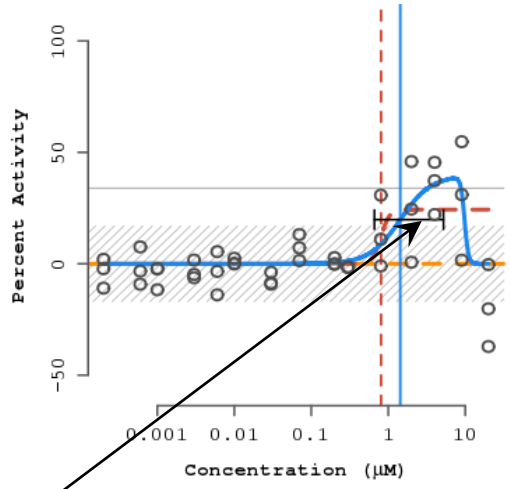


Mc7 summarizes uncertainty information for a curve-fit

- Reference vignette for description and table/fields:https://cran.r-project.org/web/packages/tcpl/vignettes/Introduction_Appendices.html#appendix-e-curve-fitting-uncertainty
- Watt, E. D. and R. S. Judson (2018). "Uncertainty quantification in ToxCast high throughput screening." PLoS One **13(7): e0196963.**
- Manuscript that examines some ways to use this and other information to filter ToxCast data is in preparation (Brown, Judson, Paul Friedman, *in prep*)

Table 17: Fields in mc7 table.

Field	Description
m4id	Level 4 ID
Aeid	Assay endpoint ID
Aenm	Assay endpoint name
Asid	Assay source ID
Acid	Assay component ID
Hit_pct	Total percent of hit calls made after 1000 bootstraps
Total_hitc	Total number of hit calls made after 1000 bootstraps
Modl_ga_min	Low bound of the 95% confidence interval for the AC_{50}
Modl_ga_max	Upper bound of the 95% confidence interval for the AC_{50}
Modl_ga_med	Median AC_{50} after 1000 bootstraps
Modl_gw_med	Median gain Hill coefficient for 1000 bootstraps
Modl_ga_delta	AC_{50} confidence interval width in log units
Cnst_pct	Percent of 1000 bootstraps that the constant model was selected as the winning model
Hill_pct	Percent of 1000 bootstraps that the Hill model was selected as the winning model
Gnls_pct	Percent of 1000 bootstraps that the gain-loss was selected as the winning model



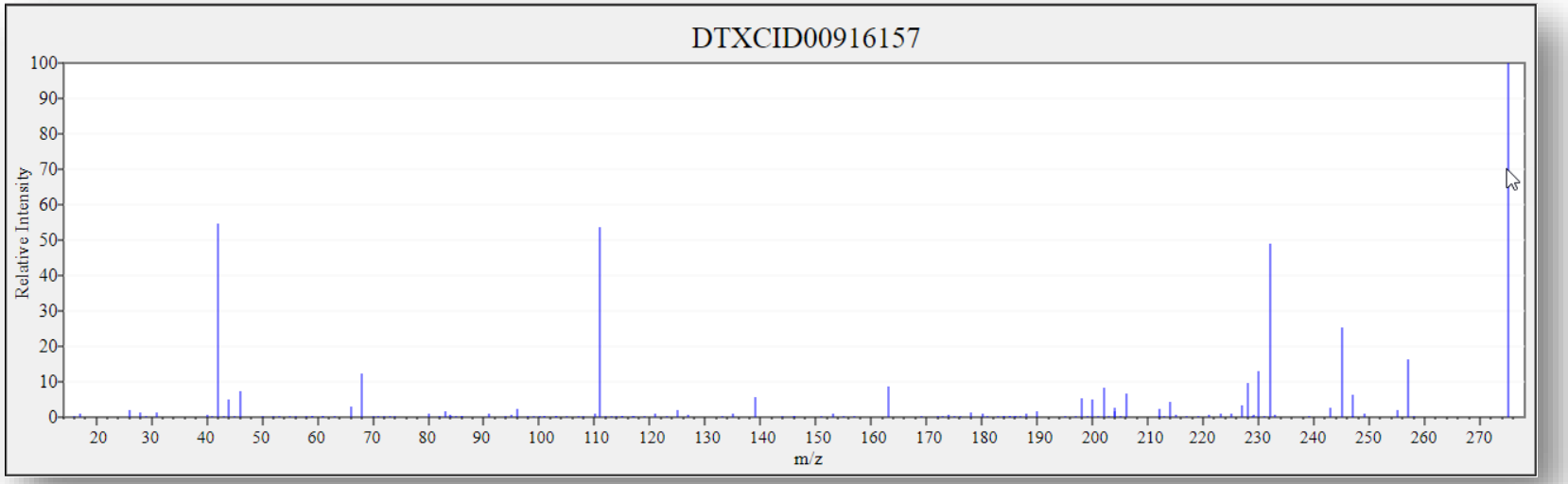
WORK IN PROGRESS

Work in Progress

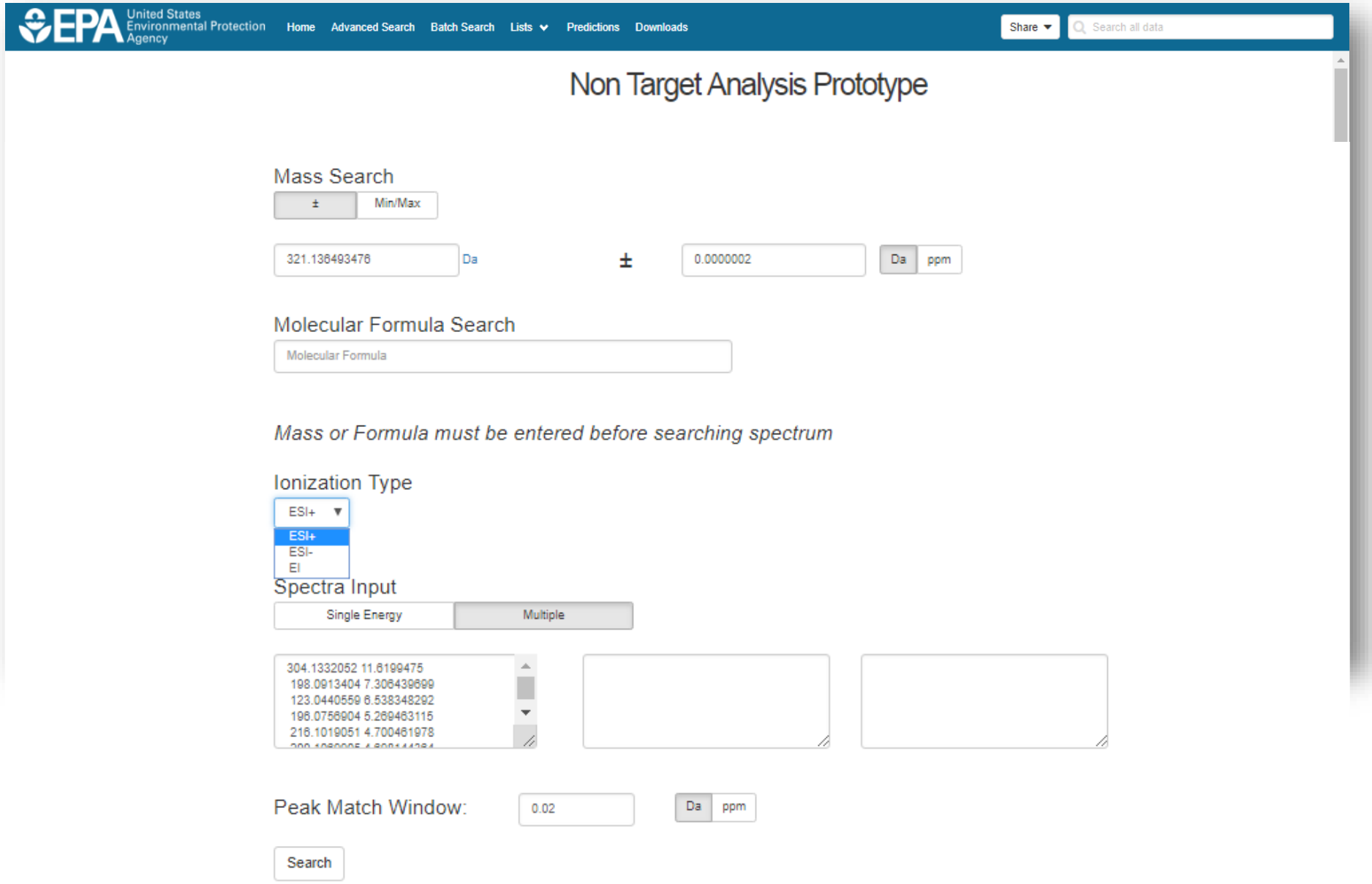
- Predicted Spectra for candidate ranking
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database

Predicted Mass Spectra

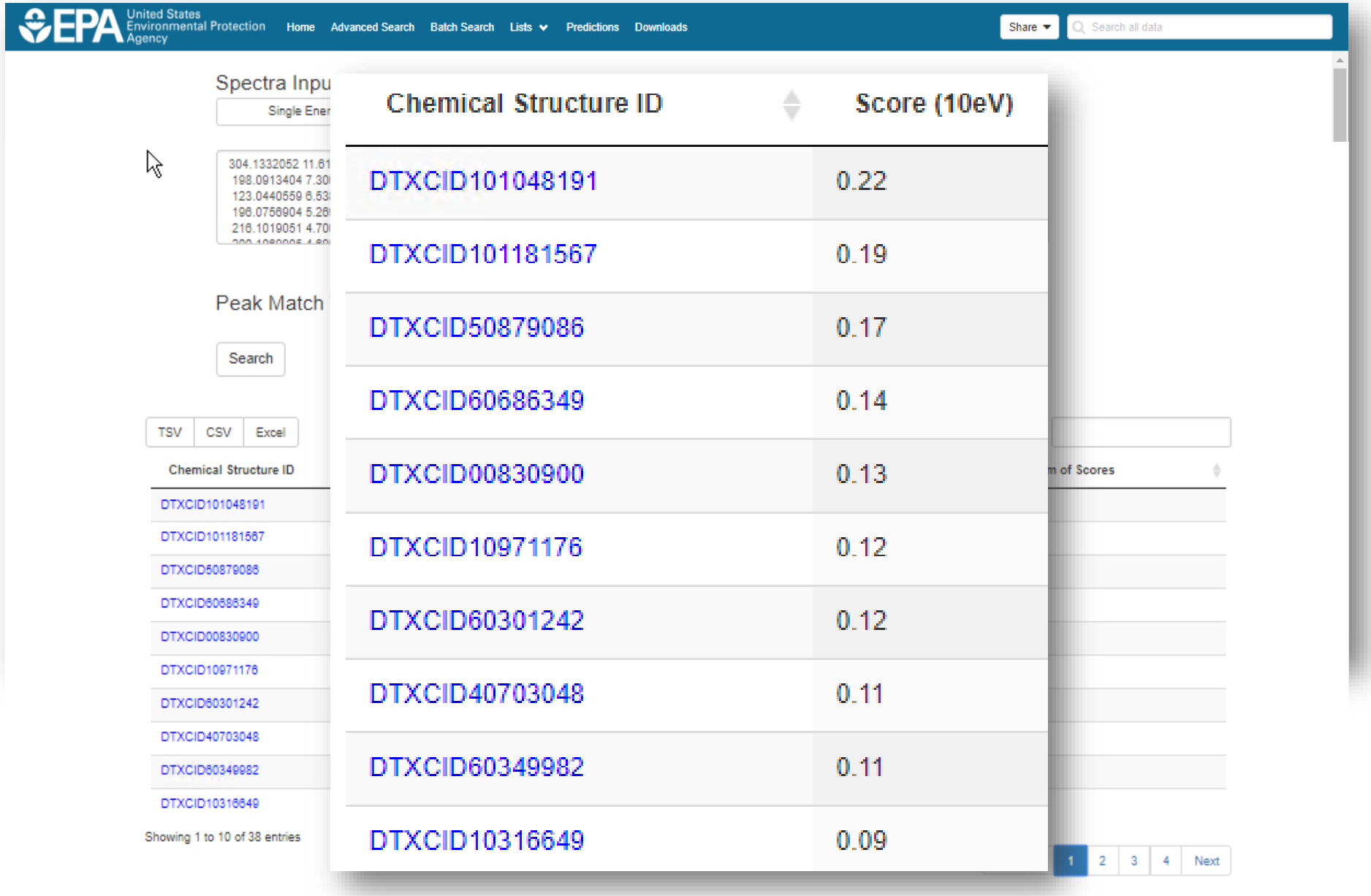
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



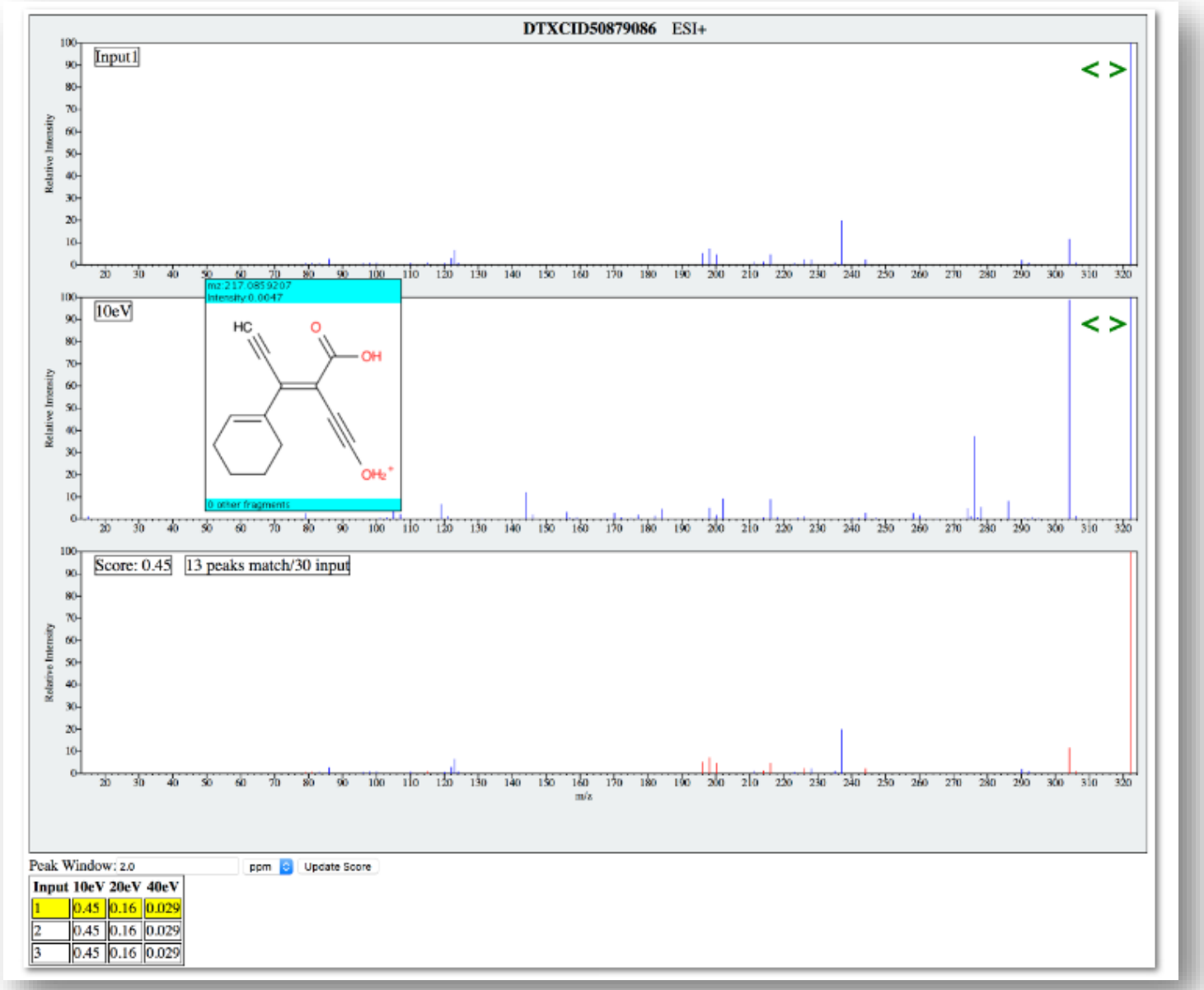
Search Expt. vs. Predicted Spectra



Search Expt. vs. Predicted Spectra



Spectral Viewer Comparison



Work in Progress

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

Including Relative Retention Times

ELSEVIER

Journal of Hazardous Materials

Volume 363, 5 February 2019, Pages 277-285

Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants

Reza Aalizadeh, Maria-Christina Nika, Nikolaos S. Thomaidis

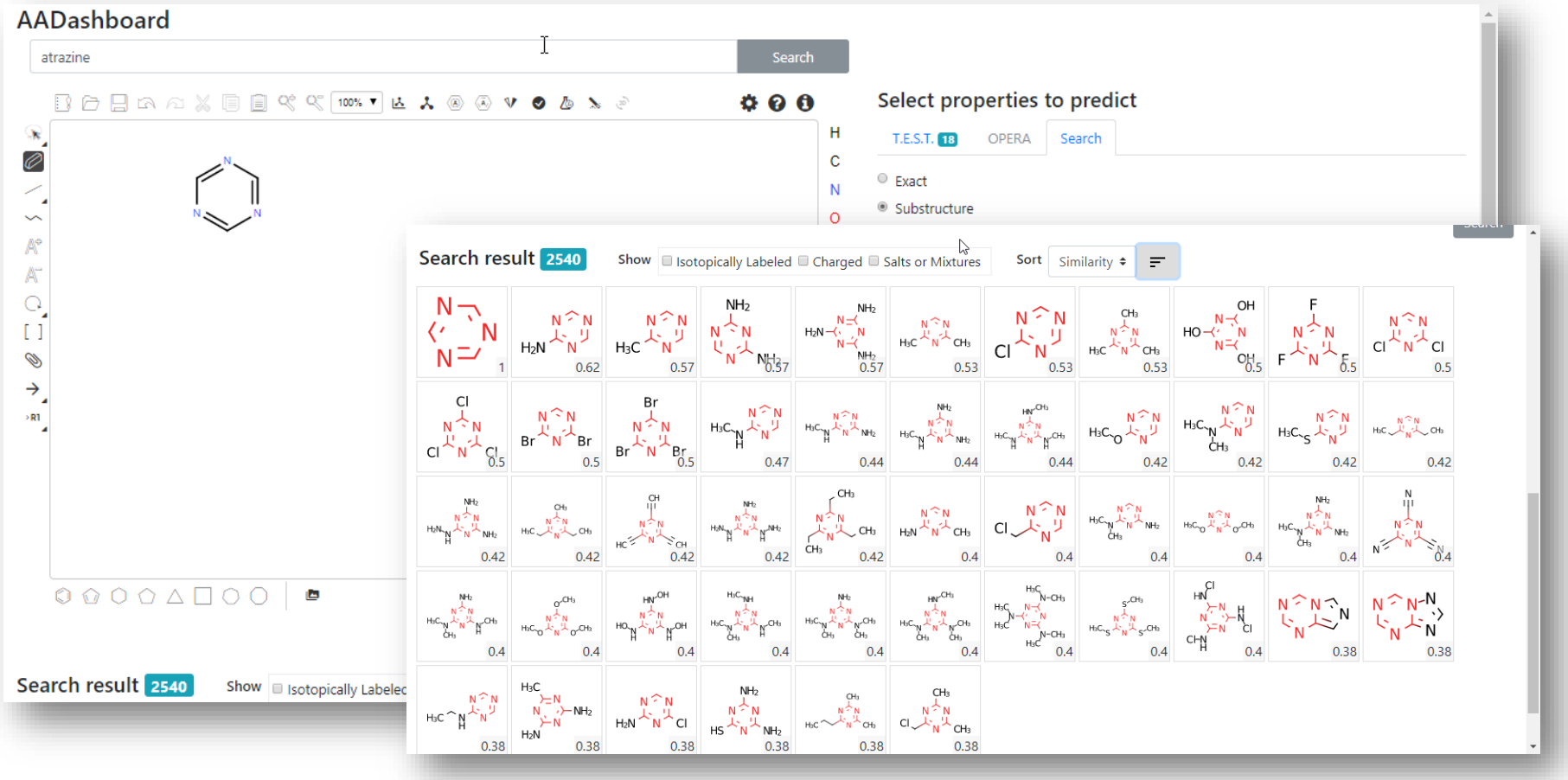
Show more

https://doi.org/10.1016/j.jhazmat.2018.09.047

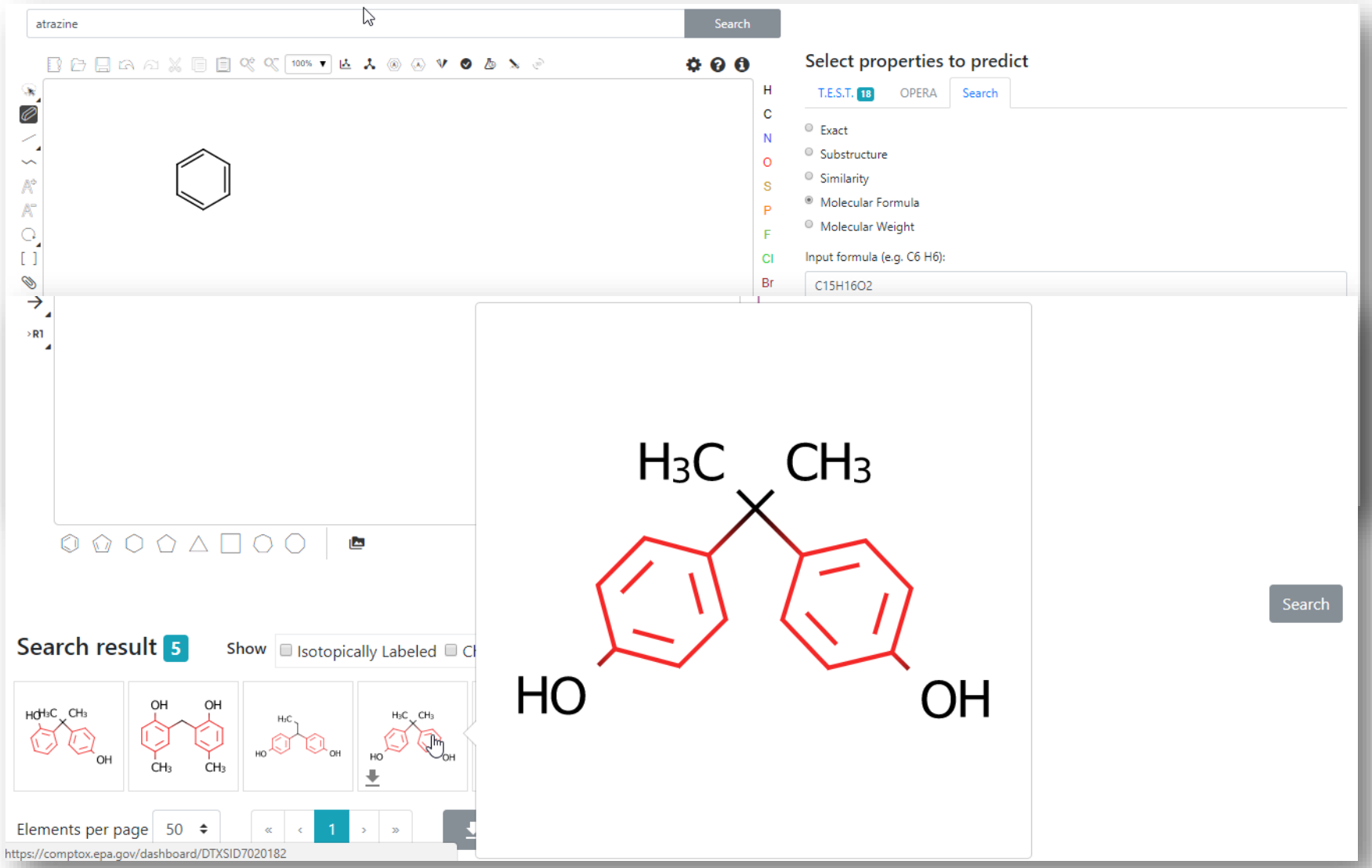
Work in Progress

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

Prototype Development



Prototype Development

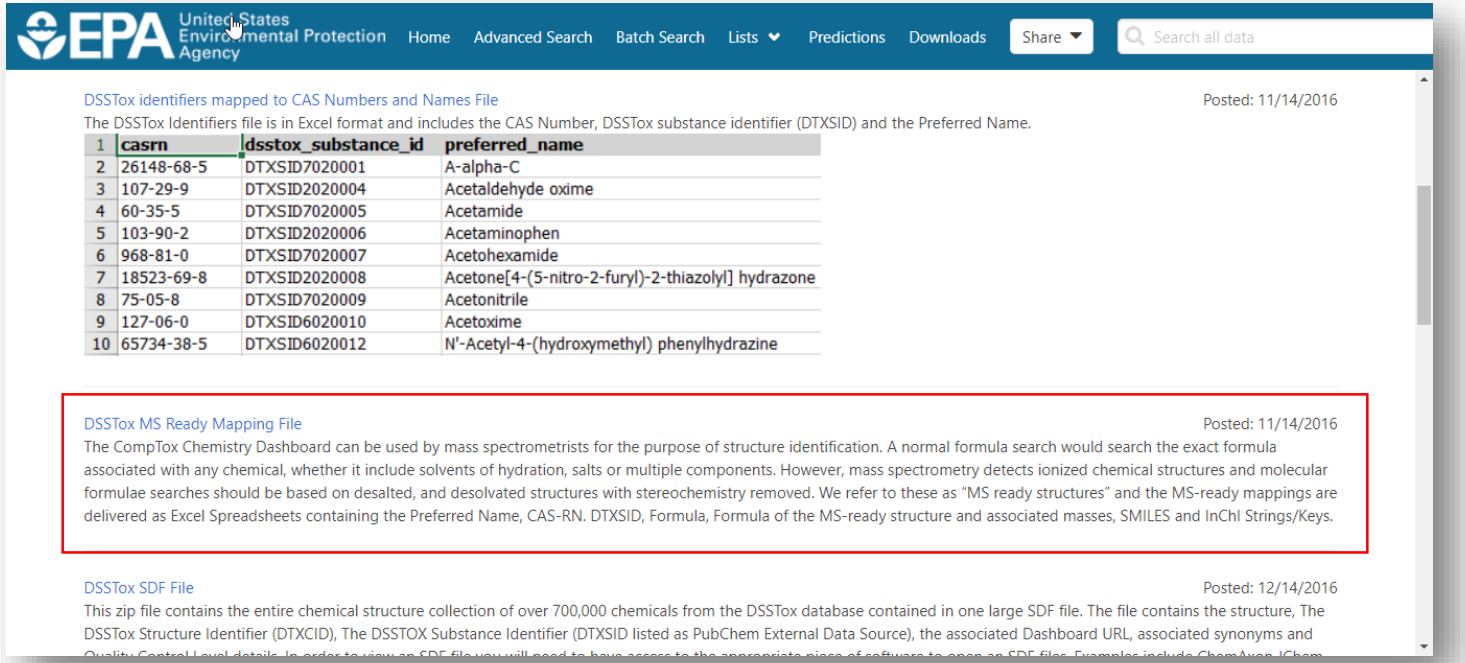


Work in Progress

- Predicted Spectra for candidate ranking
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search
- Access to API and web services for programmatic access

API services and Open Data

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



Other Work in Progress and Future Plans

- **Work in Progress**
- New manual in development
- Training “videos” will be posted to YouTube in the future
- New lists are in preparation to add to the lists page

- **Future Work**
- Integrate ToxRefDB database views
- Realtime OPERA predictions for physchem/fate and transport
- Integrate H295R model views and other models as available

Acknowledgements for v3 release

- The NCCT IT development team – led by Jeff Edwards and Jeremy Dunne
- NCCT PIs –Chris Grulke, Keith Houck, Richard Judson, Grace Patlewicz, Ann Richard, Imran Shah, John Wambaugh
- ORISE/SSC – Jason Brown, Andrew McEachran, Nathaniel Rush, Anita Simha, Mahmoud Shobair

Feedback welcome

paul-friedman.katie@epa.gov

williams.antony@epa.gov