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

Translating research into practical tools: A case study of GenRA, a new read-across tool

Antony Williams[1], George Helman[2], Jeff Edwards[1], Jeremy Dunne[1], Imran Shah[1] and Grace Patlewicz[1]

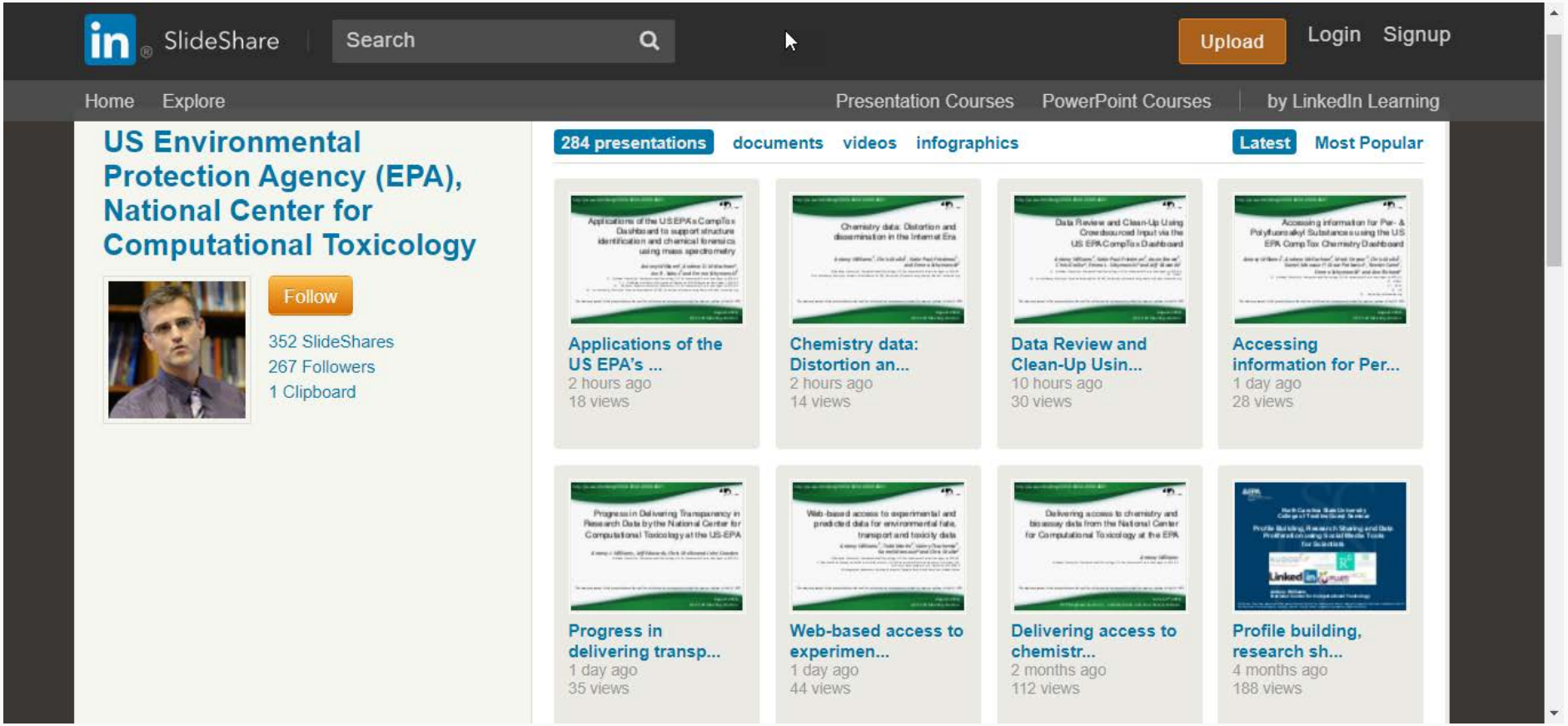
1) National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC
2) Oak Ridge Institute of Science and Education (ORISE) Research Participant, RTP, NC

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

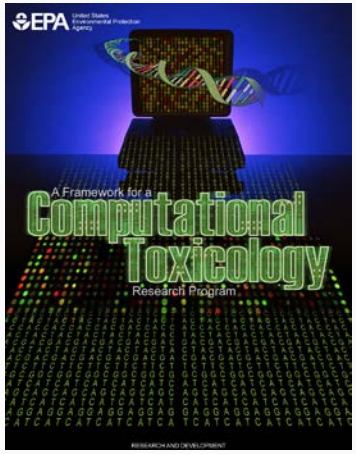
August 2018
ACS Fall Meeting, Boston

Slides Will Be Available Here

https://www.slideshare.net/AntonyWilliams/presentations



National Center for Computational Toxicology



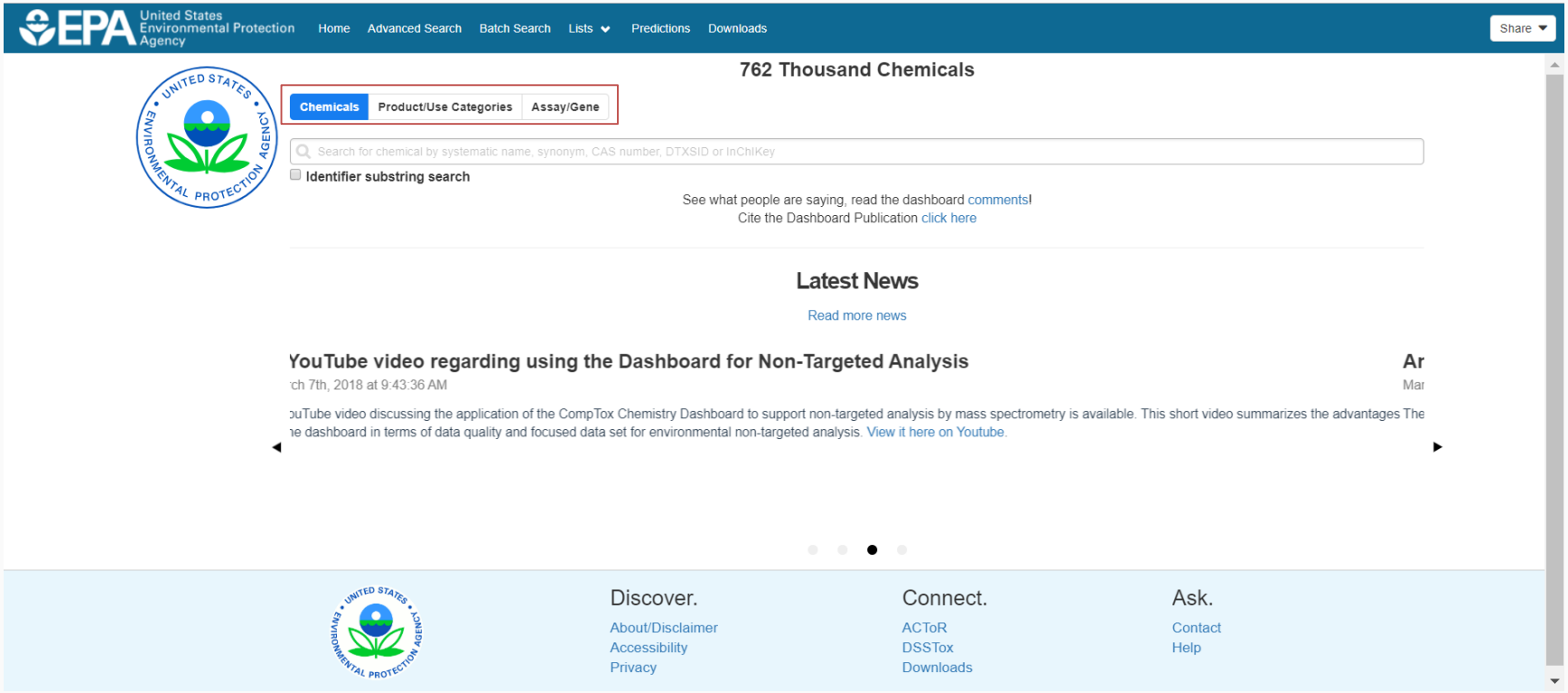
- National Center for Computational Toxicology established in 2005 to integrate:
  - High-throughput and high-content technologies
  - Modern molecular biology
  - Data mining and statistical modeling
  - Computational biology and chemistry
- Researching computational approaches to quickly evaluate the safety of chemicals for potential risk.
- Outputs: a lot of data, models, algorithms and software applications

The CompTox Dashboard
https://comptox.epa.gov/dashboard

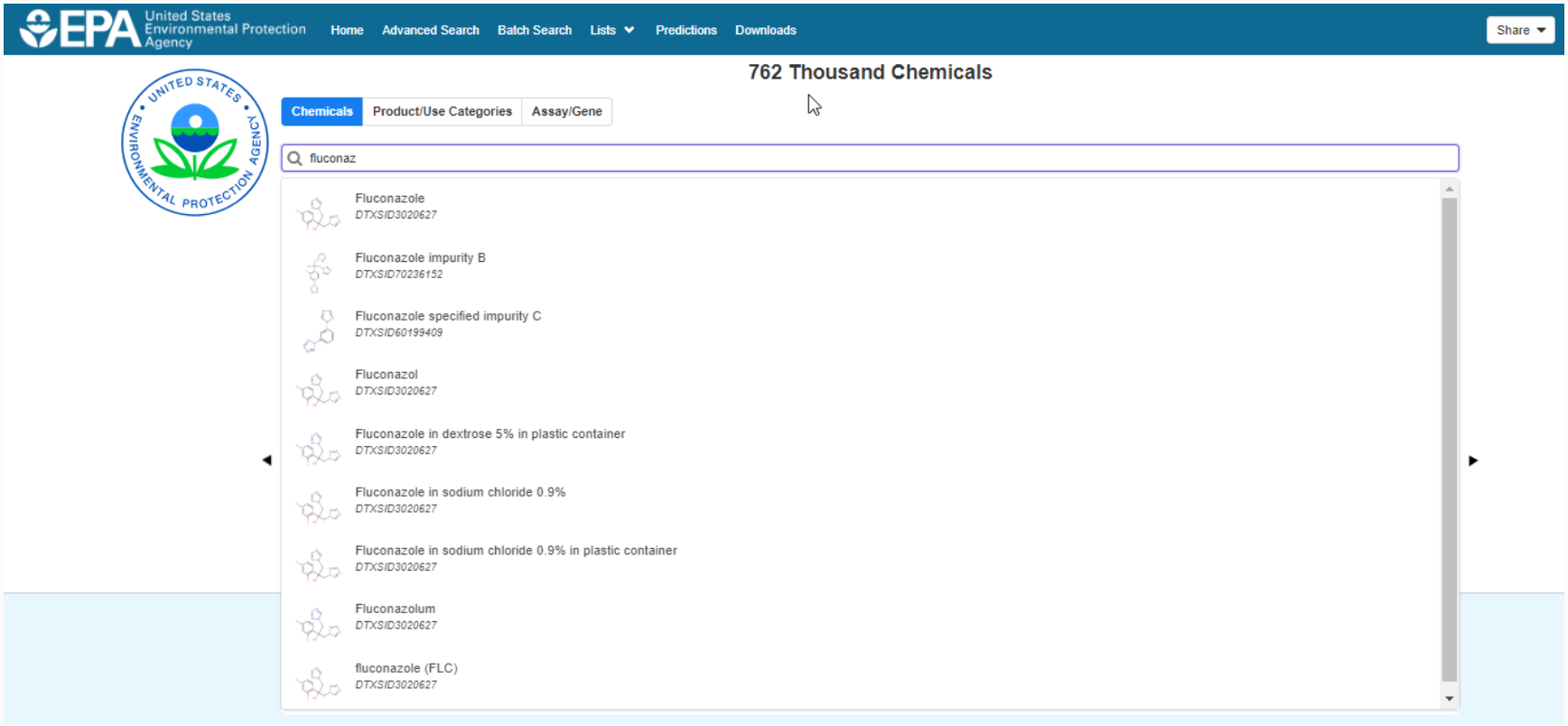
- A publicly accessible website delivering access:
  - ~762,000 chemicals with related property data
  - Searchable by chemical, product use, gene and assay
  - **Experimental** and **predicted** physicochemical property data, environmental fate and transport, and tox endpoints
  - “Bioactivity data” for the ToxCast/Tox21 project – plus derived **models**
  - **NEW** Generalized **Read-Across** (GenRA) module
  - “Batch searching” of **predicted** data for 1000s of chemicals

CompTox Dashboard

https://comptox.epa.gov/dashboard



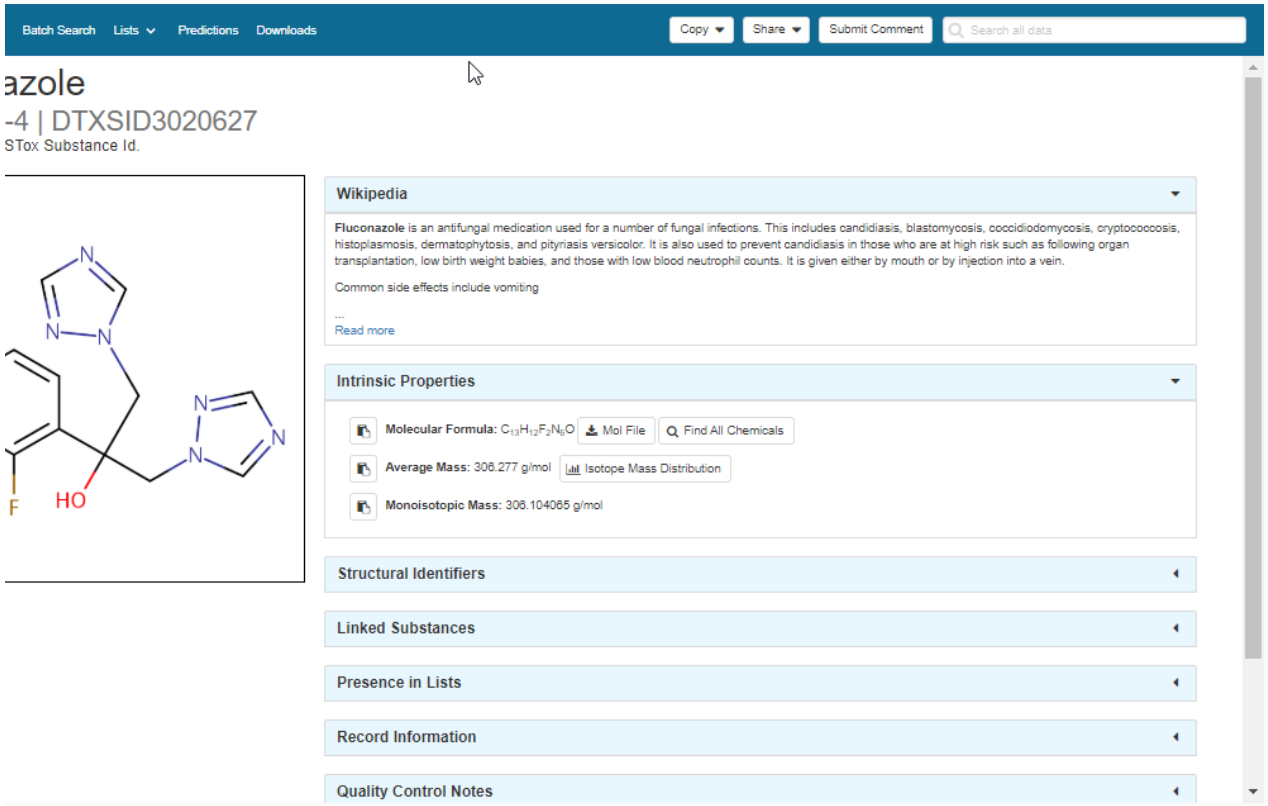
CompTox Dashboard Chemicals



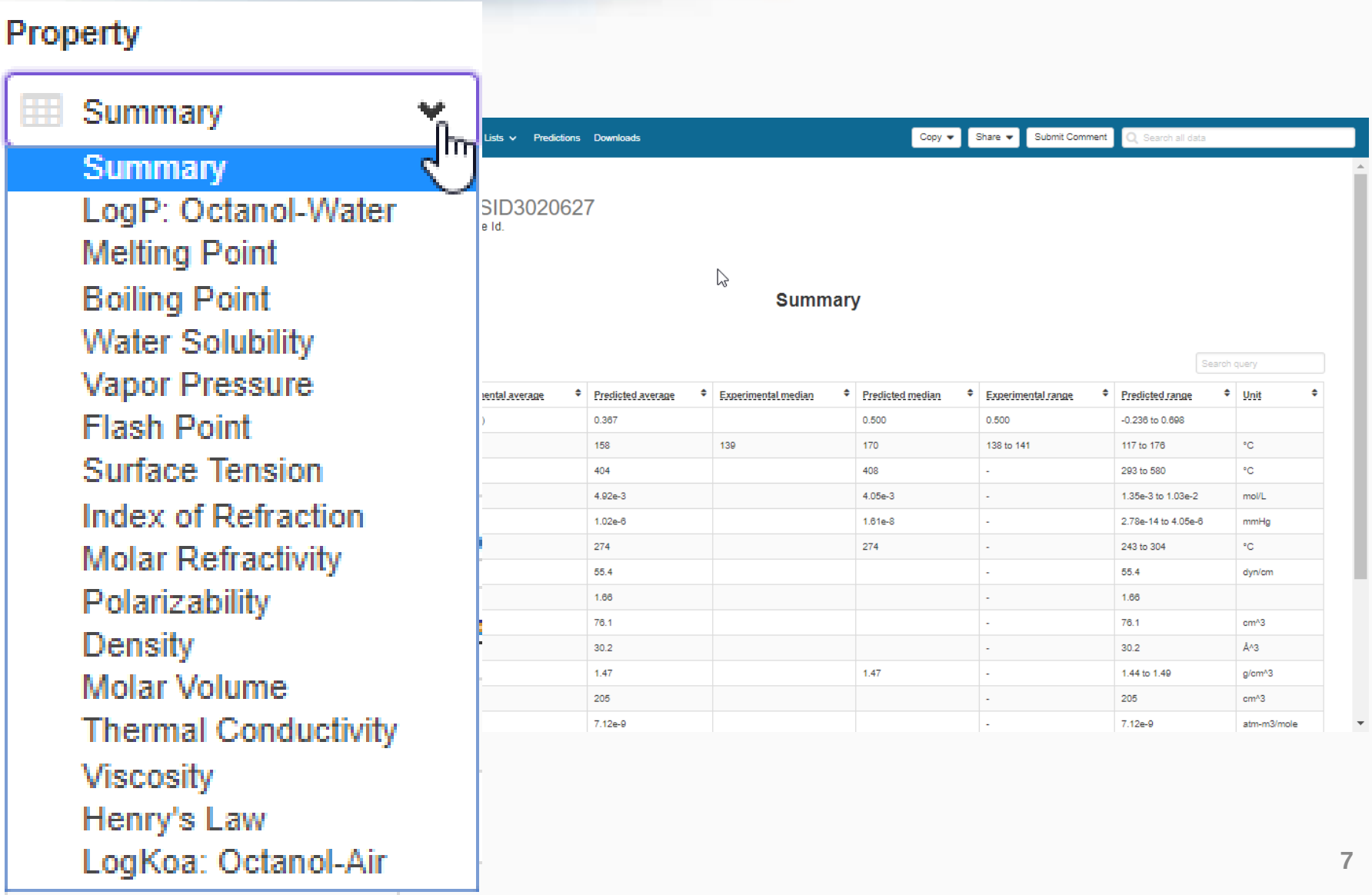
Detailed Chemical Pages

EPA United States Environmental Protection Agency

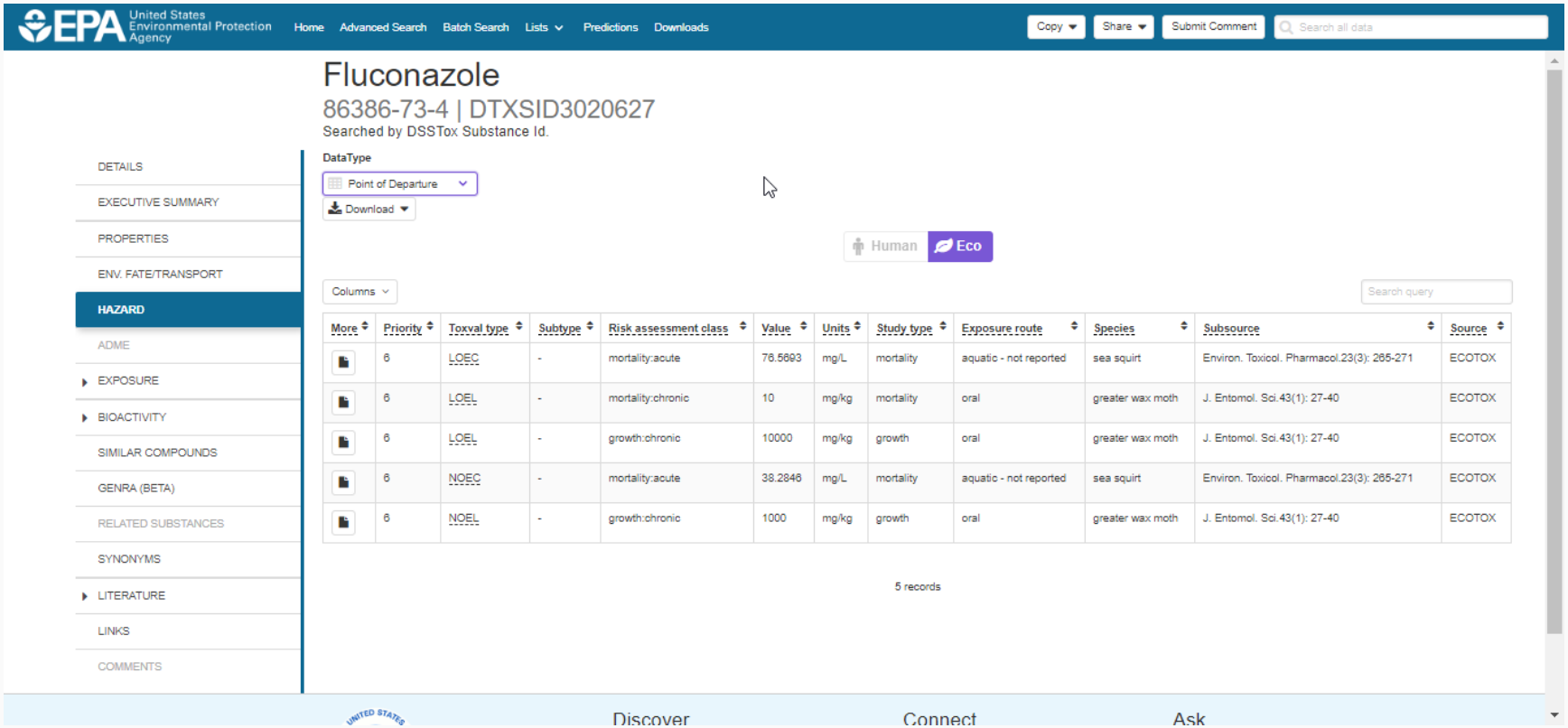
- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS



Physicochemical properties

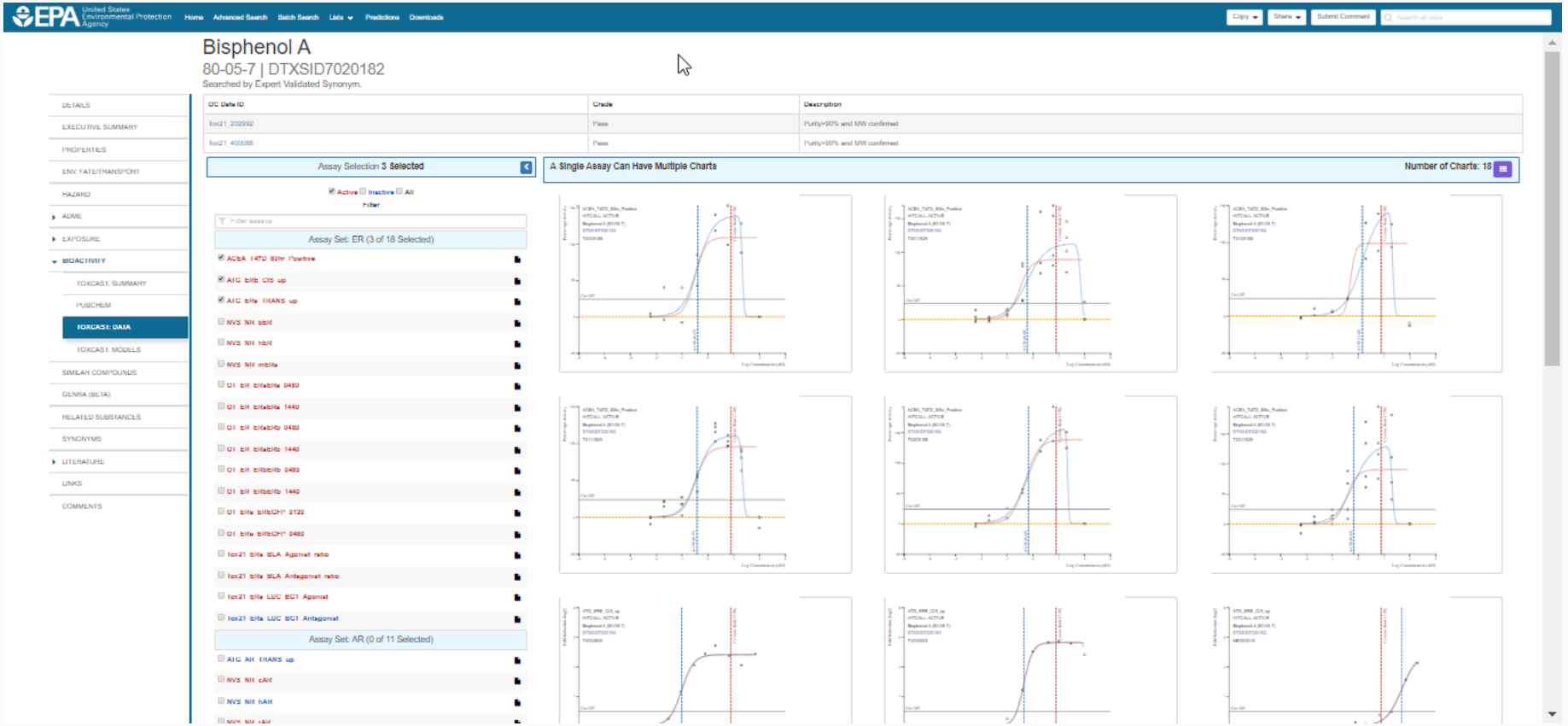


Hazard Data – Human and Eco



Bioactivity Data (ToxCast/Tox21)

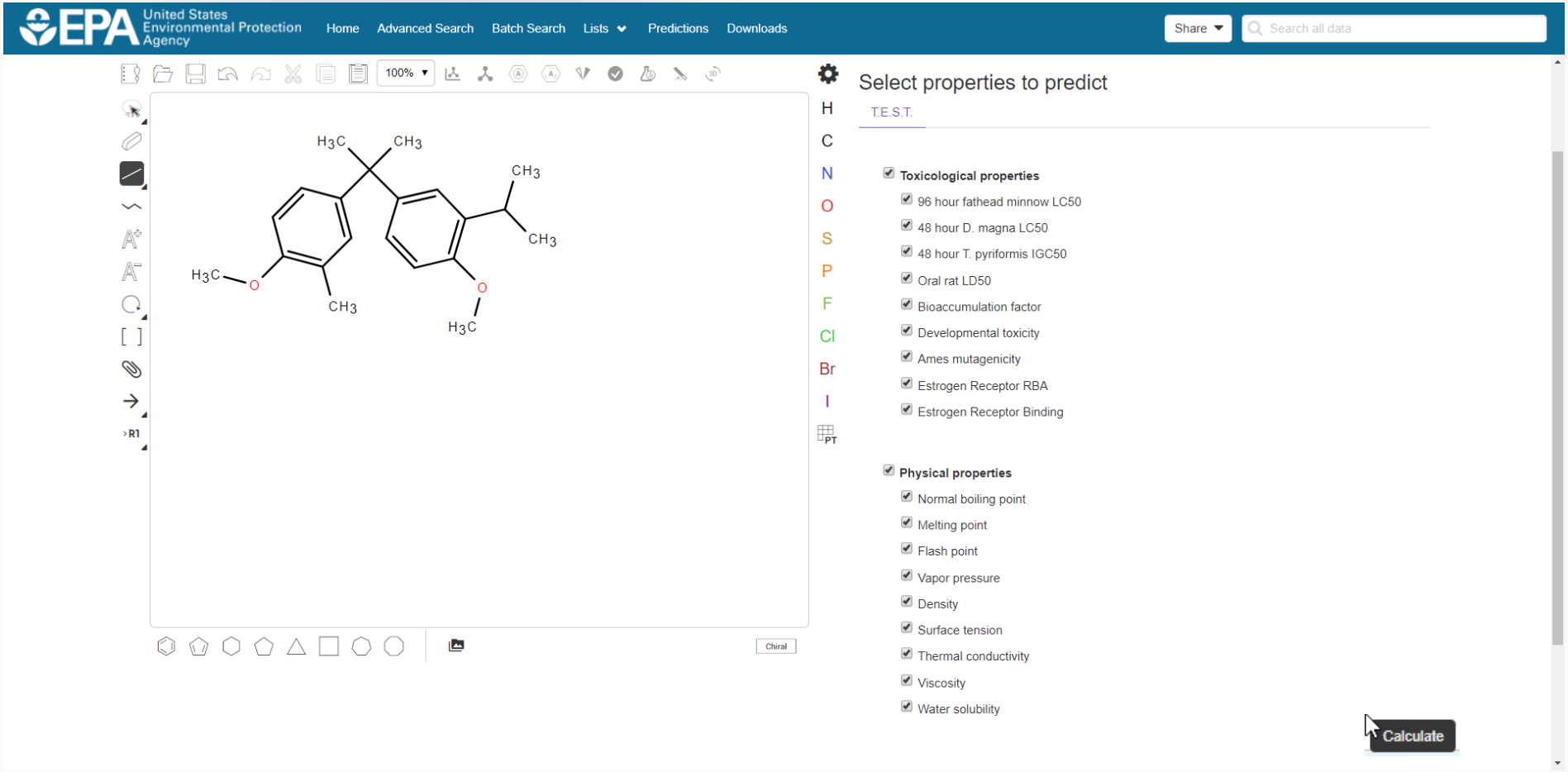
Data below for Bisphenol A



Other Dashboard Predictions

- Predictions and models expand outside of simply physicochemical and environmental fate and transport
- Examples
  - Read-across for Toxicity Endpoints
  - Quantitative Structure–Use Relationship (QSUR) models
  - High-Throughput ToxicoKinetics (HTTK)
  - Models based on high throughput bioactivity data

Real-Time Predictions



Real-Time Predictions

EPA United States Environmental Protection Agency

Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Share | Search all data

Provider: T.E.S.T.

Download Summary

Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50		6.051 -Log10(mol/L) 0.278 mg/L	5.678 -Log10(mol/L) 0.656 mg/L	5.572 -Log10(mol/L) 0.836 mg/L	5.908 -Log10(mol/L) 0.386 mg/L	7.047 -Log10(mol/L) 0.028 mg/L
48 hour D. magna LC50		5.591 -Log10(mol/L) 0.802 mg/L	5.548 -Log10(mol/L) 0.884 mg/L	6.169 -Log10(mol/L) 0.212 mg/L	5.518 -Log10(mol/L) 0.948 mg/L	5.128 -Log10(mol/L) 2.329 mg/L
48 hour T. pyriformis IGC50		5.590 -Log10(mol/L) 0.804 mg/L	6.390 -Log10(mol/L) 0.127 mg/L		5.588 -Log10(mol/L) 0.806 mg/L	4.790 -Log10(mol/L) 5.068 mg/L
Oral rat LD50		2.400 -Log10(mol/kg) 1243.951 mg/kg	2.232 -Log10(mol/kg) 1829.942 mg/kg			2.568 -Log10(mol/kg) 845.609 mg/kg
Bioaccumulation factor		3.066 Log10 1164.438	3.090 Log10 1230.849	2.717 Log10 521.420	3.257 Log10 1806.262	3.200 Log10 1585.959
Developmental toxicity		true	true	true		true
Ames mutagenicity		false	false			false
Estrogen Receptor RBA		-0.710 Log10 0.195	-1.692 Log10 0.020	-1.515 Log10 0.031		1.077 Log10 11.931
Estrogen Receptor Binding		false	false	false		true
Normal boiling point		345.2 °C	306.6 °C		408.2 °C	320.7 °C
Melting point		74.3 °C	63.8 °C		41.0 °C	118.2 °C
Flash point		161.7 °C	143.5 °C		152.7 °C	188.9 °C
Vapor pressure		-5.955 Log10(mmHg) 1.109*10^-6 mmHg	-5.534 Log10(mmHg) 2.925*10^-6 mmHg		-5.903 Log10(mmHg) 1.249*10^-6 mmHg	-6.428 Log10(mmHg) 3.735*10^-7 mmHg
Density		0.959 g/cm³	0.977 g/cm³		0.843 g/cm³	1.057 g/cm³

Real-Time Predictions

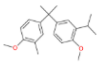
Predicted Fathead minnow LC50 (96 hr) for O(C1=CC=C(C=C1C)C(C2=CC=C(OC)C(=C2)C(C)C)(C)C)C from Consensus method

Prediction results

Endpoint	Experimental value	Predicted value
Fathead minnow LC_{50} (96 hr) -Log10(mol/L)	N/A	6.05
Fathead minnow LC_{50} (96 hr) mg/L	N/A	0.28

Individual Predictions

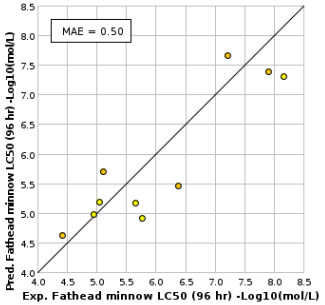
Method	Predicted value -Log10(mol/L)
Hierarchical clustering	5.68
Single model	5.57
Group contribution	5.91
Nearest neighbor	7.05



Predictions for the test chemical and for the most similar chemicals in the external test set

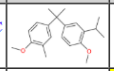
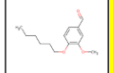
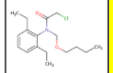
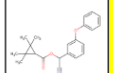
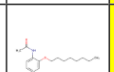
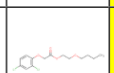
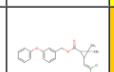
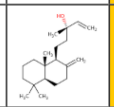
If the predicted value matches the experimental values for similar chemicals in the test set (and the similar chemicals were predicte

Prediction results (colors defined in table below)



Chemicals	MAE*
Entire set	0.55
Similarity coefficient ≥ 0.5	0.50

*Mean absolute error in -Log10(mol/L)

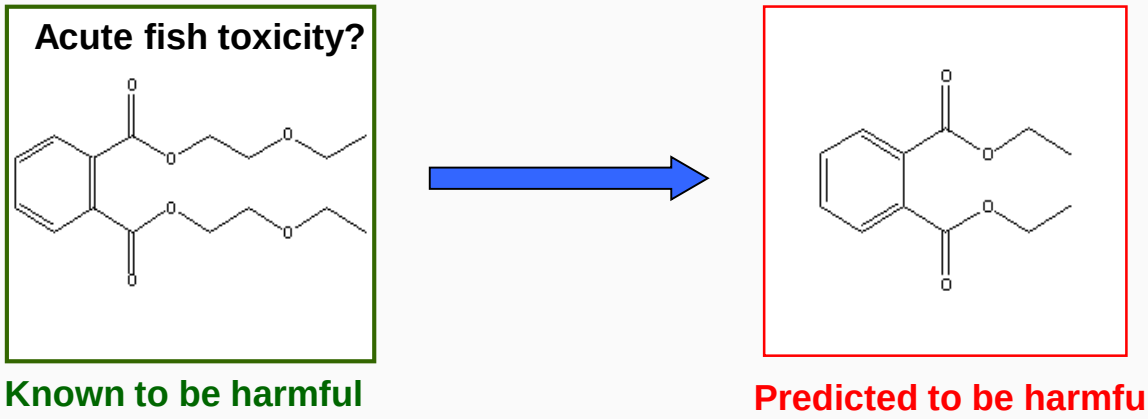
CAS	Structure	Similarity Coefficient	Experimental value -Log10(mol/L)	Predicted value -Log10(mol/L)
O(C1=CC=C(C=C1C)C(C2=CC=C(OC)C(=C2)C(C)C)(C)C)C (test chemical)			N/A	6.05
61096-84-2		0.74	4.95	4.97
23184-66-9		0.73	5.65	5.17
39515-41-8		0.72	8.17	7.32
55792-61-5		0.71	5.77	4.92
1929-73-3		0.71	5.05	5.19
52645-53-1		0.70	7.21	7.67
596-85-0		0.70	6.38	5.47

Definitions: Read-Across

- Known information on the property of a substance (source) is used to make a prediction of the same property for another substance (target) that is considered “similar”

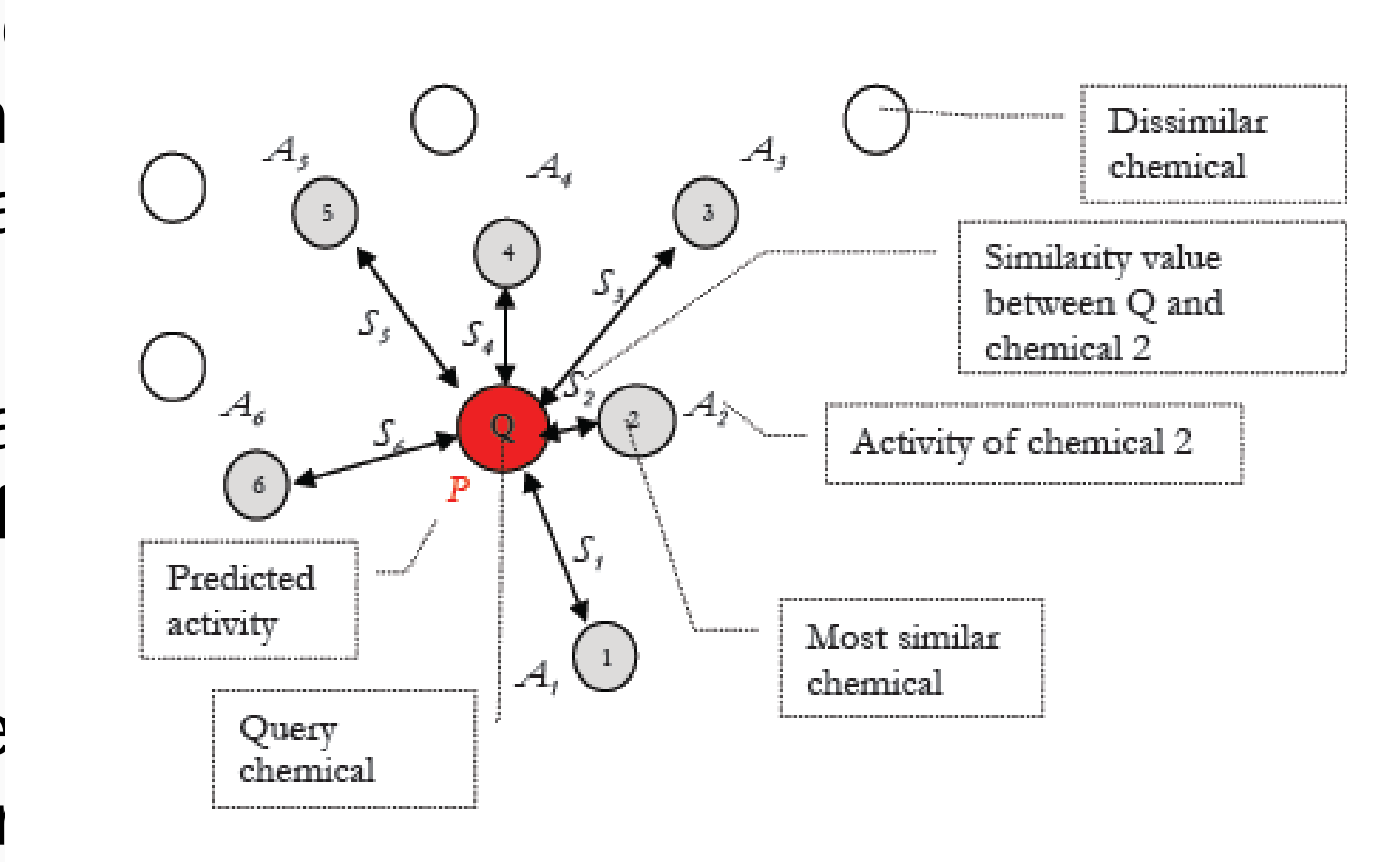
	Source chemical	Target chemical
Property	● →	○

● Reliable data
○ Missing data

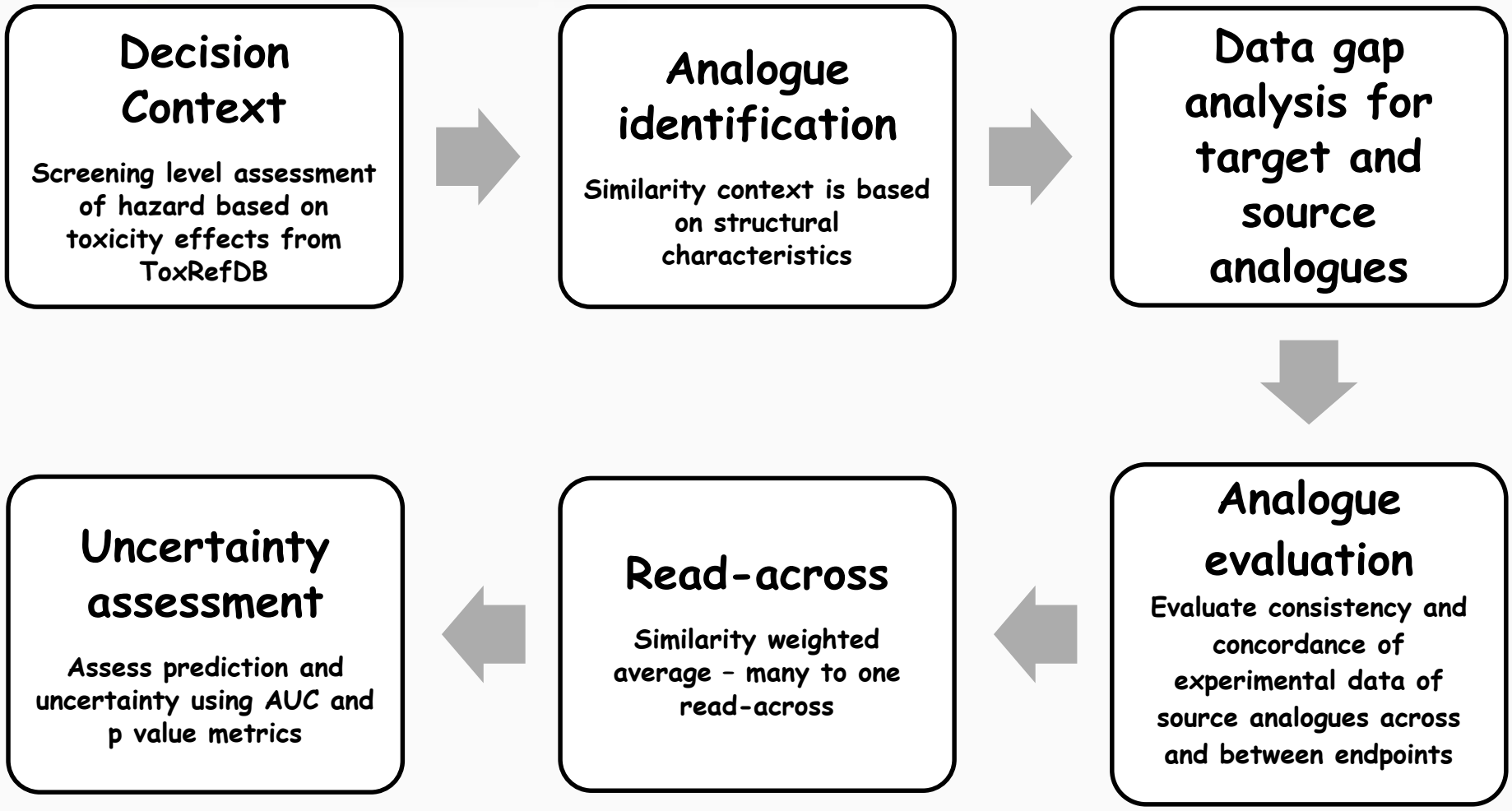


GenRA (Generalised Read-Across)

- Pre vity of n or bioa
- Goa data perf
- The e for within specific study outcomes to be established

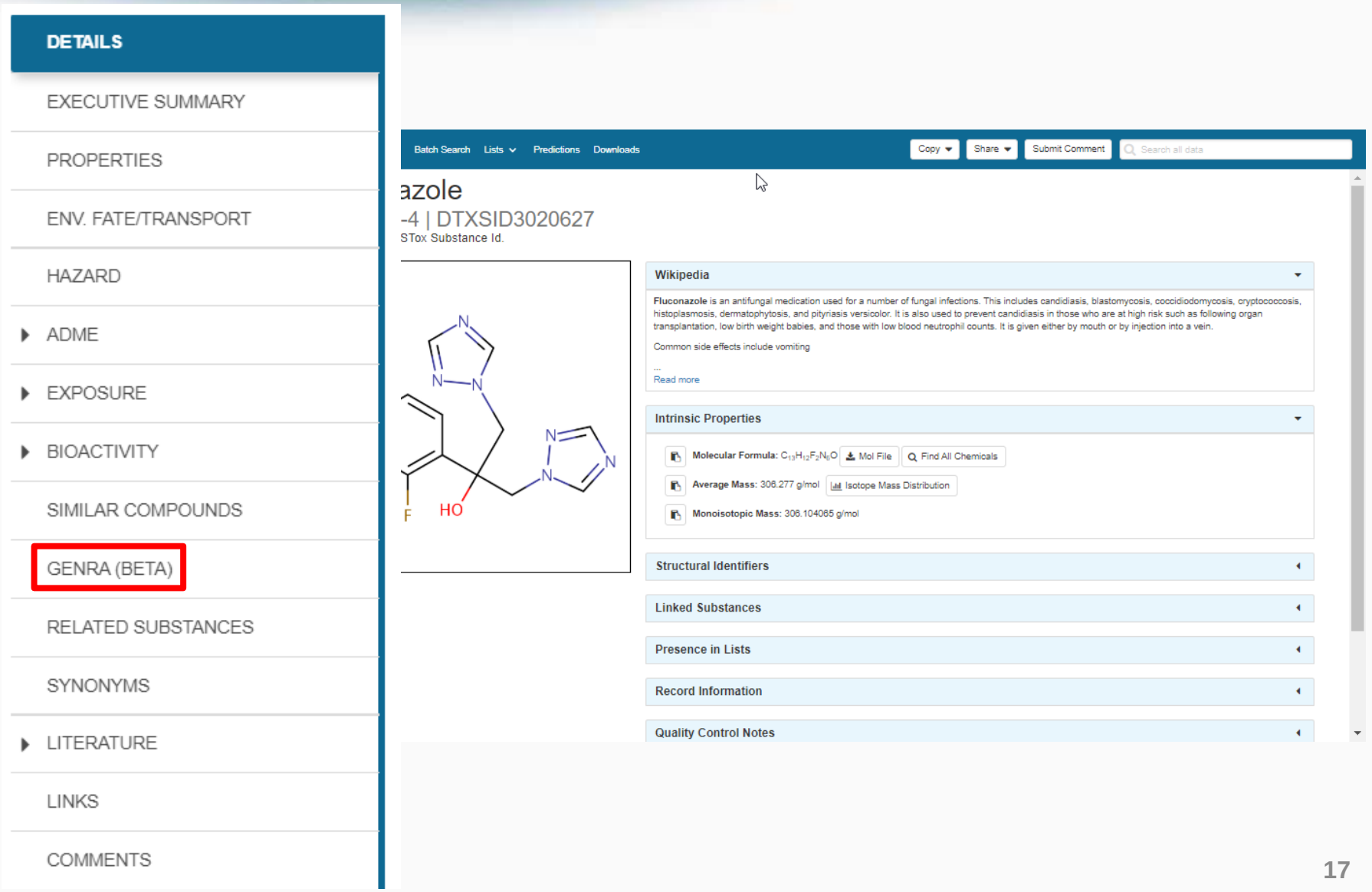


Read-across workflow in GenRA

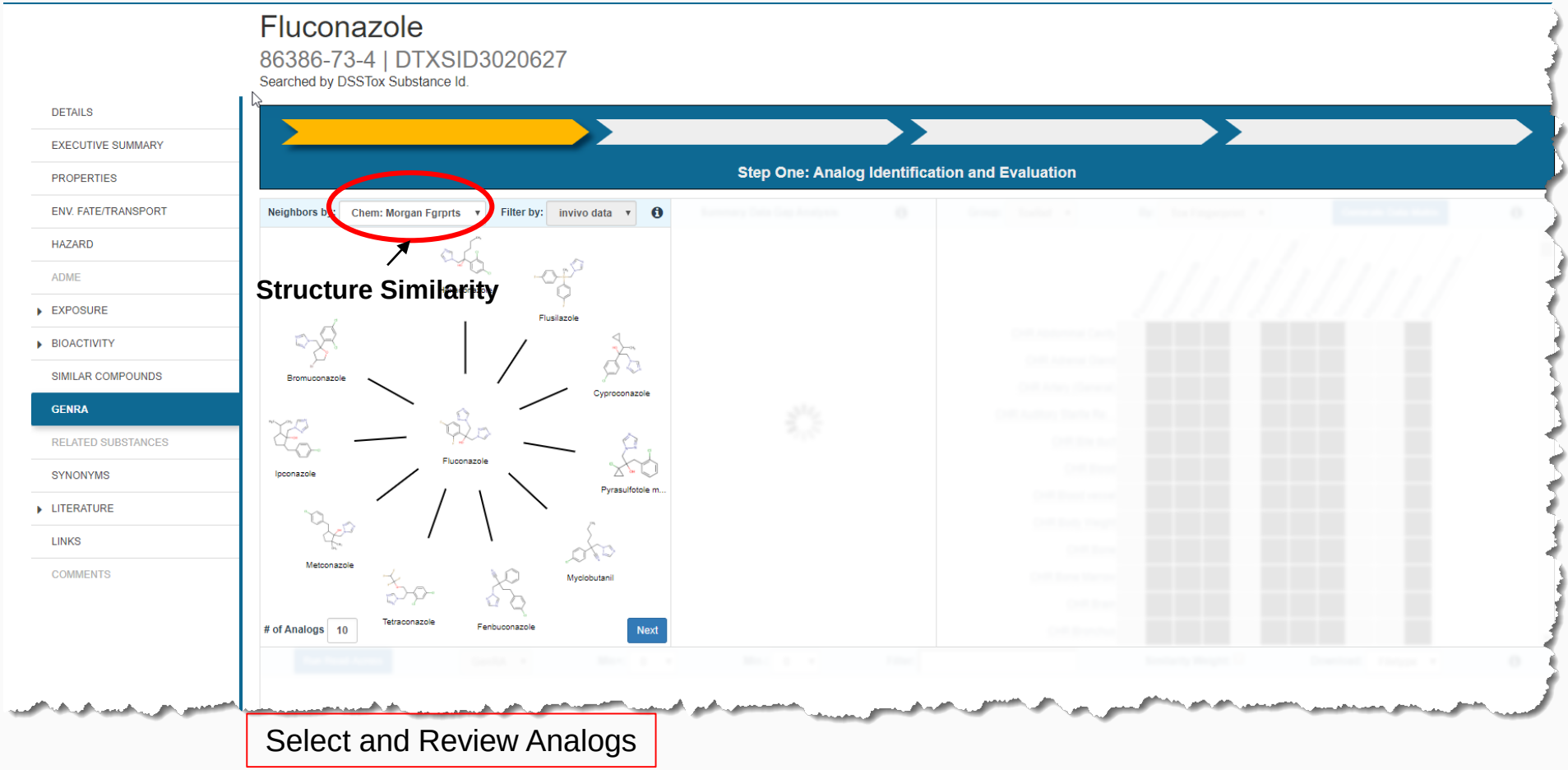


GenRA (Generalised Read-Across)

EPA United States Environmental Protection Agency

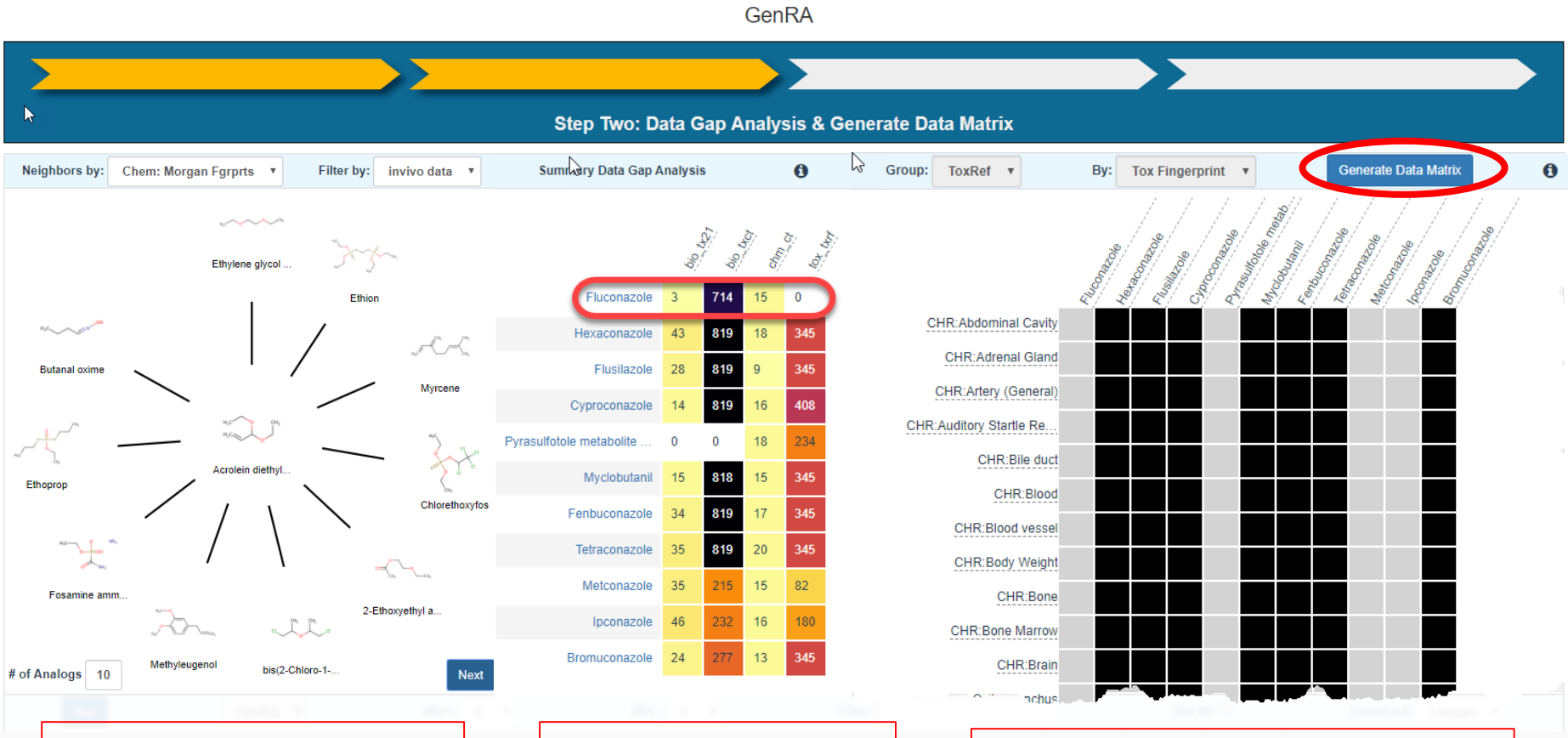


GenRA (Generalised Read-Across)



Select and Review Analogs

GenRA (Generalised Read-Across)

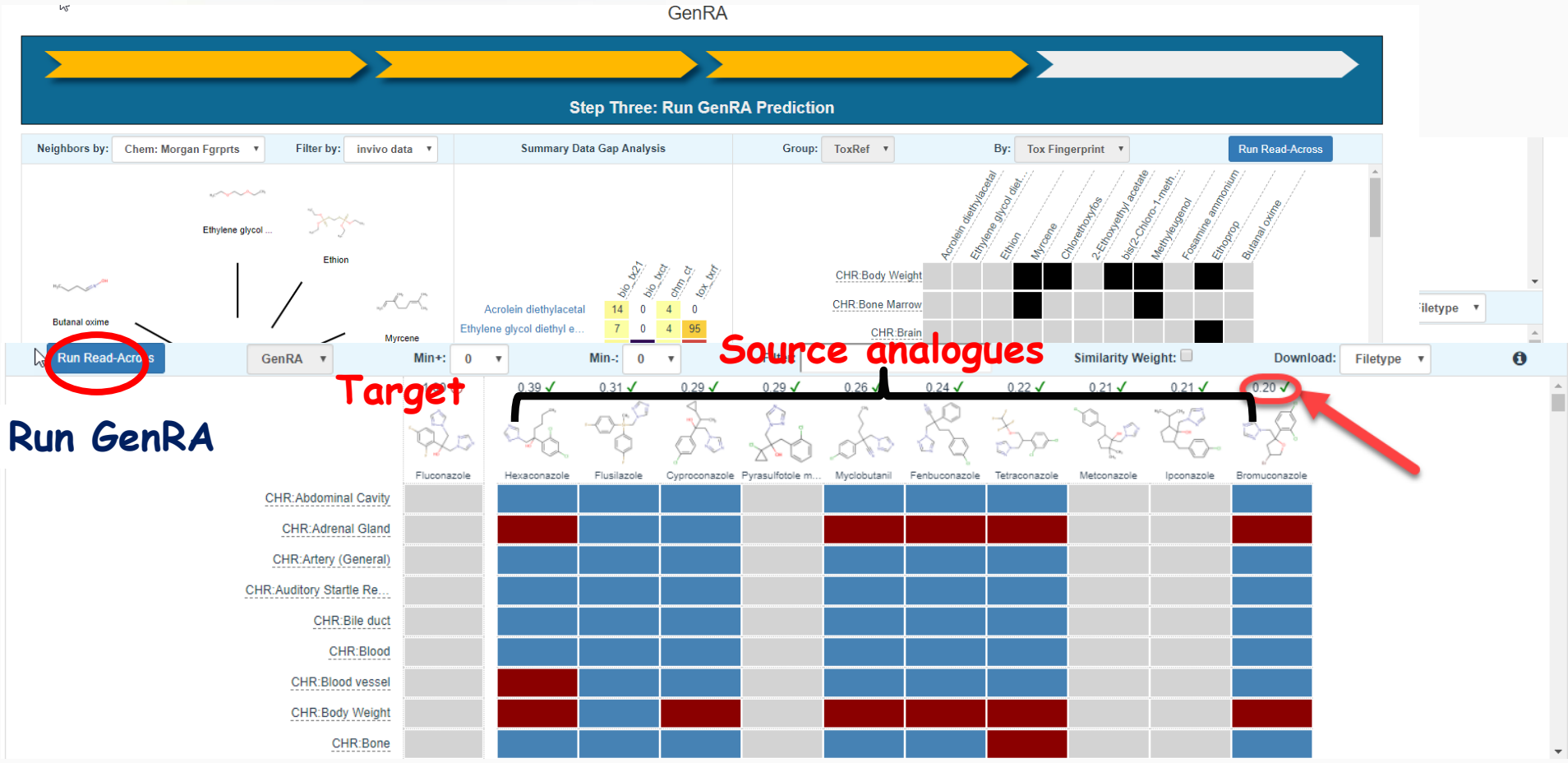


Select and Review Analogs

Review Available Data

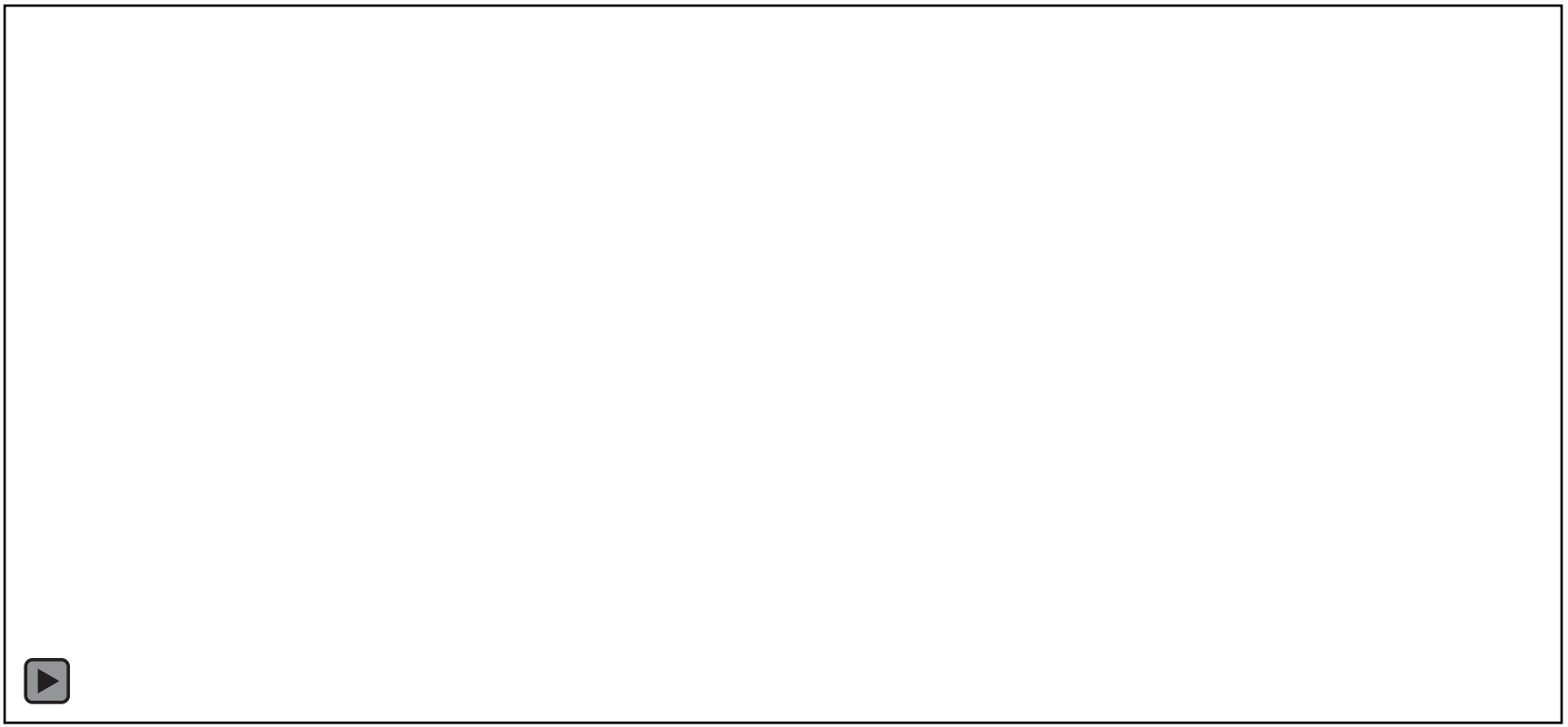
Fingerprint indicating available data

GenRA (Generalised Read-Across)



Red : Toxicity effects.
Blue: No Toxicity effects
Grey : Absence of data

Demonstration



GenRA is one of multiple Read-Across Tools available

Tool	AIM	ToxMatch	AMBIT	OECD Toolbox	CBRA	ToxRead	GenRA
Analogue identification	X	X	X	X	X	X	X
Analogue Evaluation	NA	X	X by other tools available	X	X	X For Ames & BCF	NA
Data gap analysis	NA	X	X Data matrix can be exported	X Data matrix viewable	NA	NA	X Data matrix can be exported
Data gap filling	NA	X	User driven	X	X	X	X
Uncertainty assessment	NA	NA	NA	X	NA	NA	X
Availability	Free	Free	Free	Free	Free	Free	Free

Related Publications

Chemical Research in Toxicology

Article

Cite This: Chem. Res. Toxicol. 2017, 30, 2046-2059 pubs.acs.org/crt

Predicting Organ Toxicity Using *in Vitro* Bioactivity Data and Chemical Structure

Jie Liu,[‡,§] Grace Patlewicz,[†] Antony J. Williams,[†] Russell S. Thomas,[†] and Imran Shah*[,†]

[†]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, Durham, North Carolina 27711, United States

[‡]Department of Information Science, University of Arkansas at Little Rock, Arkansas 72204, United States

[§]Oak Ridge Institute for Science Education, National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, Durham, North Carolina 27711, United States

ELSEVIER

Regulatory Toxicology and Pharmacology

Volume 79, August 2016, Pages 12-24

Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information

Imran Shah [a], Jie Liu [b, c], Richard S. Judson [a], Russell S. Thomas [a], Grace Patlewicz [a]

ELSEVIER

Computational Toxicology

Available online 23 July 2018

In Press, Corrected Proof

Extending the Generalised Read-Across approach (GenRA): A systematic analysis of the impact of physicochemical property information on read-across performance

George Helman [a, b], Imran Shah [b], Grace Patlewicz [b]

ELSEVIER

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com

Journal Cover Image

Navigating through the minefield of read-across frameworks: A commentary perspective

Grace Patlewicz[a, *], Mark T.D. Cronin[b], George Helman[a, c], Jason C. Lambert[d], Lucina E. Lizarraga[d], Imran Shah[a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency (US EPA), 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA

[b] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

[c] Oak Ridge Institute for Science and Education (ORISE), 1299 Bethel Valley Road, Oak Ridge, TN 37830, USA

[d] National Center for Evaluation Assessment (NCEA), US Environmental Protection Agency (US EPA), 26 West Martin Luther King Dr, Cincinnati, OH 45268, USA

PROOF

Conclusions

- The CompTox Dashboard delivers experimental and predicted data for physchem, environ. fate and transport

- A new **Read-Across** module, GenRA, is now available

- Real time predictions are also possible – coming soon pKa and logD predictions

Acknowledgments

National Center Comp. Tox.

- Imran Shah
- George Helman
- Prachi Pradeep
- Tony Williams
- Jeff Edwards
- Jeremy Dunne
- NCCT Development team
- Chris Grulke
- Reeder Sams
- Katie-Paul Friedman
- Rusty Thomas

National Center for Environ. Assessment

- Jason Lambert
- Lucy Lizarraga

- Mark Cronin LJMU

Contact

Antony Williams
US EPA Office of Research and Development
National Center for Computational Toxicology (NCCT)
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821