

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

*Antony Williams¹, George Helman², Jeff Edwards¹, Jeremy Dunne¹,
Imran Shah¹ and Grace Patlewicz¹*

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>



Search



Upload

Login Signup

Home Explore

Presentation Courses

PowerPoint Courses

by LinkedIn Learning

US Environmental Protection Agency (EPA), National Center for Computational Toxicology



Follow

352 SlideShares
267 Followers
1 Clipboard

284 presentations

documents

videos

infographics

Latest

Most Popular



Applications of the US EPA's ...
2 hours ago
18 views



Chemistry data: Distortion an...
2 hours ago
14 views



Data Review and Clean-Up Usin...
10 hours ago
30 views



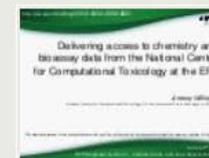
Accessing information for Per...
1 day ago
28 views



Progress in delivering transp...
1 day ago
35 views



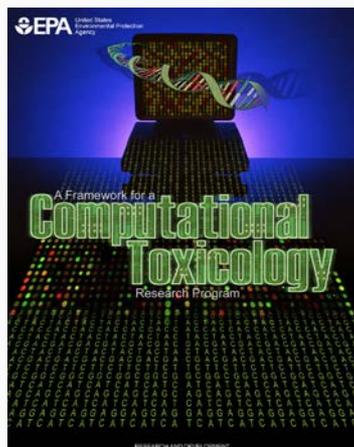
Web-based access to experimen...
1 day ago
44 views



Delivering access to chemistr...
2 months ago
112 views



Profile building, research sh...
4 months ago
188 views



- 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

- 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



Chemicals Product/Use Categories Assay/Gene

762 Thousand Chemicals

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](#)

YouTube video regarding using the Dashboard for Non-Targeted Analysis

March 7th, 2018 at 9:43:36 AM

Ar
Mar

A YouTube video discussing the application of the CompTox Chemistry Dashboard to support non-targeted analysis by mass spectrometry is available. This short video summarizes the advantages of the dashboard in terms of data quality and focused data set for environmental non-targeted analysis. [View it here on Youtube.](#)



Discover.

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

Connect.

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

Ask.

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

CompTox Dashboard Chemicals



762 Thousand Chemicals

Chemicals Product/Use Categories Assay/Gene

fluconaz

-  Fluconazole
DTXSID3020627
-  Fluconazole impurity B
DTXSID70236152
-  Fluconazole specified impurity C
DTXSID60199409
-  Fluconazol
DTXSID3020627
-  Fluconazole in dextrose 5% in plastic container
DTXSID3020627
-  Fluconazole in sodium chloride 0.9%
DTXSID3020627
-  Fluconazole in sodium chloride 0.9% in plastic container
DTXSID3020627
-  Fluconazolium
DTXSID3020627
-  fluconazole (FLC)
DTXSID3020627

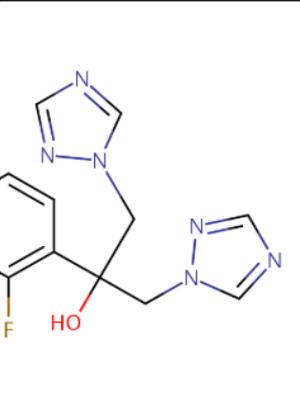
Detailed Chemical Pages

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

Batch Search Lists Predictions Downloads Copy Share Submit Comment Search all data

azole

-4 | DTXSID3020627
STox Substance Id.



Wikipedia

Fluconazole is an antifungal medication used for a number of fungal infections. This includes candidiasis, blastomycosis, coccidioidomycosis, cryptococcosis, histoplasmosis, dermatophytosis, and pityriasis versicolor. It is also used to prevent candidiasis in those who are at high risk such as following organ transplantation, low birth weight babies, and those with low blood neutrophil counts. It is given either by mouth or by injection into a vein. Common side effects include vomiting

...
[Read more](#)

Intrinsic Properties

Molecular Formula: $C_{12}H_{12}F_2N_4O$ [Mol File](#) [Find All Chemicals](#)

Average Mass: 306.277 g/mol [Isotope Mass Distribution](#)

Monoisotopic Mass: 306.104065 g/mol

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

Physicochemical properties

Property

Summary

Summary

LogP: Octanol-Water

Melting Point

Boiling Point

Water Solubility

Vapor Pressure

Flash Point

Surface Tension

Index of Refraction

Molar Refractivity

Polarizability

Density

Molar Volume

Thermal Conductivity

Viscosity

Henry's Law

LogKoa: Octanol-Air



SID3020627
e Id.

Lists Predictions Downloads Copy Share Submit Comment Search all data

Summary

Search query

Experimental average	Predicted average	Experimental median	Predicted median	Experimental range	Predicted range	Unit
	0.387		0.500	0.500	-0.238 to 0.698	
	158	139	170	138 to 141	117 to 176	°C
	404		408	-	293 to 580	°C
	4.92e-3		4.05e-3	-	1.35e-3 to 1.03e-2	mol/L
	1.02e-6		1.61e-8	-	2.78e-14 to 4.05e-6	mmHg
	274		274	-	243 to 304	°C
	55.4			-	55.4	dyn/cm
	1.68			-	1.68	
	76.1			-	76.1	cm³
	30.2			-	30.2	Å³
	1.47		1.47	-	1.44 to 1.49	g/cm³
	205			-	205	cm³
	7.12e-9			-	7.12e-9	atm-m³/mole

Hazard Data – Human and Eco

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

Fluconazole

86386-73-4 | DTXSID3020627
Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

LITERATURE

LINKS

COMMENTS

Data Type: Point of Departure

Download

Human Eco

Columns Search query

More	Priority	Toxval type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	6	LOEC	-	mortality:acute	76.5893	mg/L	mortality	aquatic - not reported	sea squirt	Environ. Toxicol. Pharmacol.23(3): 265-271	ECOTOX
	6	LOEL	-	mortality:chronic	10	mg/kg	mortality	oral	greater wax moth	J. Entomol. Sci.43(1): 27-40	ECOTOX
	6	LOEL	-	growth:chronic	10000	mg/kg	growth	oral	greater wax moth	J. Entomol. Sci.43(1): 27-40	ECOTOX
	6	NOEC	-	mortality:acute	38.2846	mg/L	mortality	aquatic - not reported	sea squirt	Environ. Toxicol. Pharmacol.23(3): 265-271	ECOTOX
	6	NOEL	-	growth:chronic	1000	mg/kg	growth	oral	greater wax moth	J. Entomol. Sci.43(1): 27-40	ECOTOX

5 records

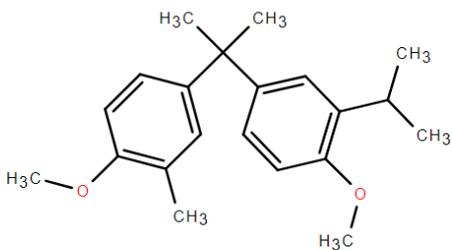
UNITED STATES Discover Connect Ask

- 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



100%



Chiral



Select properties to predict

H

TEST.

C

N

O

S

P

F

Cl

Br

I

PT

Toxicological properties

- 96 hour fathead minnow LC50
- 48 hour D. magna LC50
- 48 hour T. pyriformis IGC50
- Oral rat LD50
- Bioaccumulation factor
- Developmental toxicity
- Ames mutagenicity
- Estrogen Receptor RBA
- Estrogen Receptor Binding

Physical properties

- Normal boiling point
- Melting point
- Flash point
- Vapor pressure
- Density
- Surface tension
- Thermal conductivity
- Viscosity
- Water solubility

Calculate

Real-Time Predictions



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/cm ³	0.977 g/cm ³		0.843 g/cm ³	1.057 g/cm ³

Real-Time Predictions

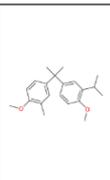
Predicted Fathead minnow LC50 (96 hr) for O(C1=CC=C(C=C1)C(C2=CC=C(OC)C(=C2)C(C)C)C)C from Consensus method

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

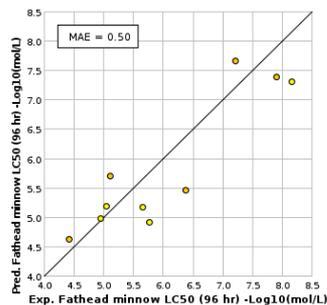
Method	Predicted value -Log10(mol/L)
Hierarchical clustering	5.68
Single model	5.57
Group contribution	5.91
Nearest neighbor	7.05



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 predicted)

Prediction results (colors defined in table below)



Chemicals	MAE*
Entire set	0.55
Similarity coefficient ≥ 0.5	0.50

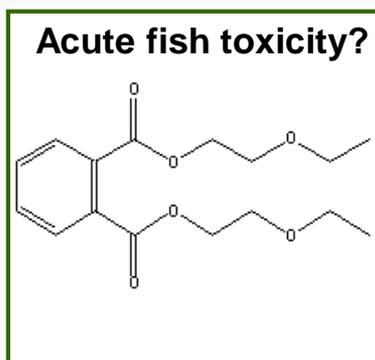
*Mean absolute error in -Log10(mol/L)

CAS	Structure	Similarity Coefficient	Experimental value -Log10(mol/L)	Predicted value -Log10(mol/L)
<chem>O(C1=CC=C(C=C1)C(C2=CC=C(OC)C(=C2)C(C)C)C)C</chem> (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
52643-53-1		0.70	7.21	7.67
596-85-0		0.70	6.38	5.47

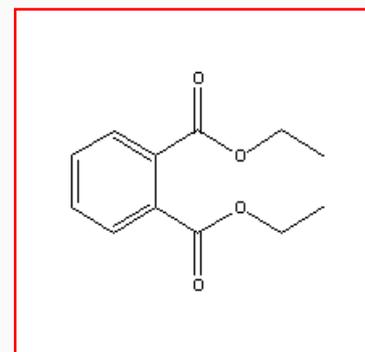
Definitions: Read-Across

- 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	●	○	<ul style="list-style-type: none"> ● Reliable data ○ Missing data



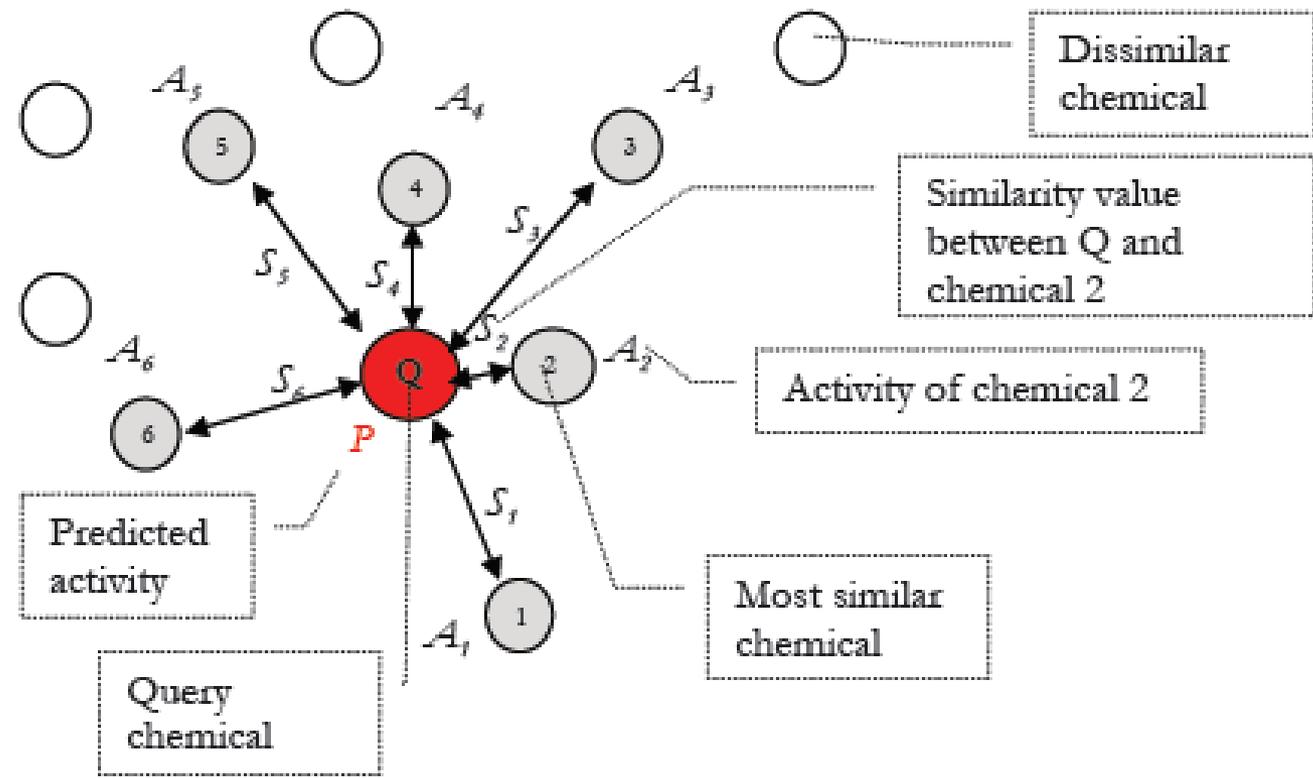
Known to be harmful



Predicted to be harmful

GenRA (Generalised Read-Across)

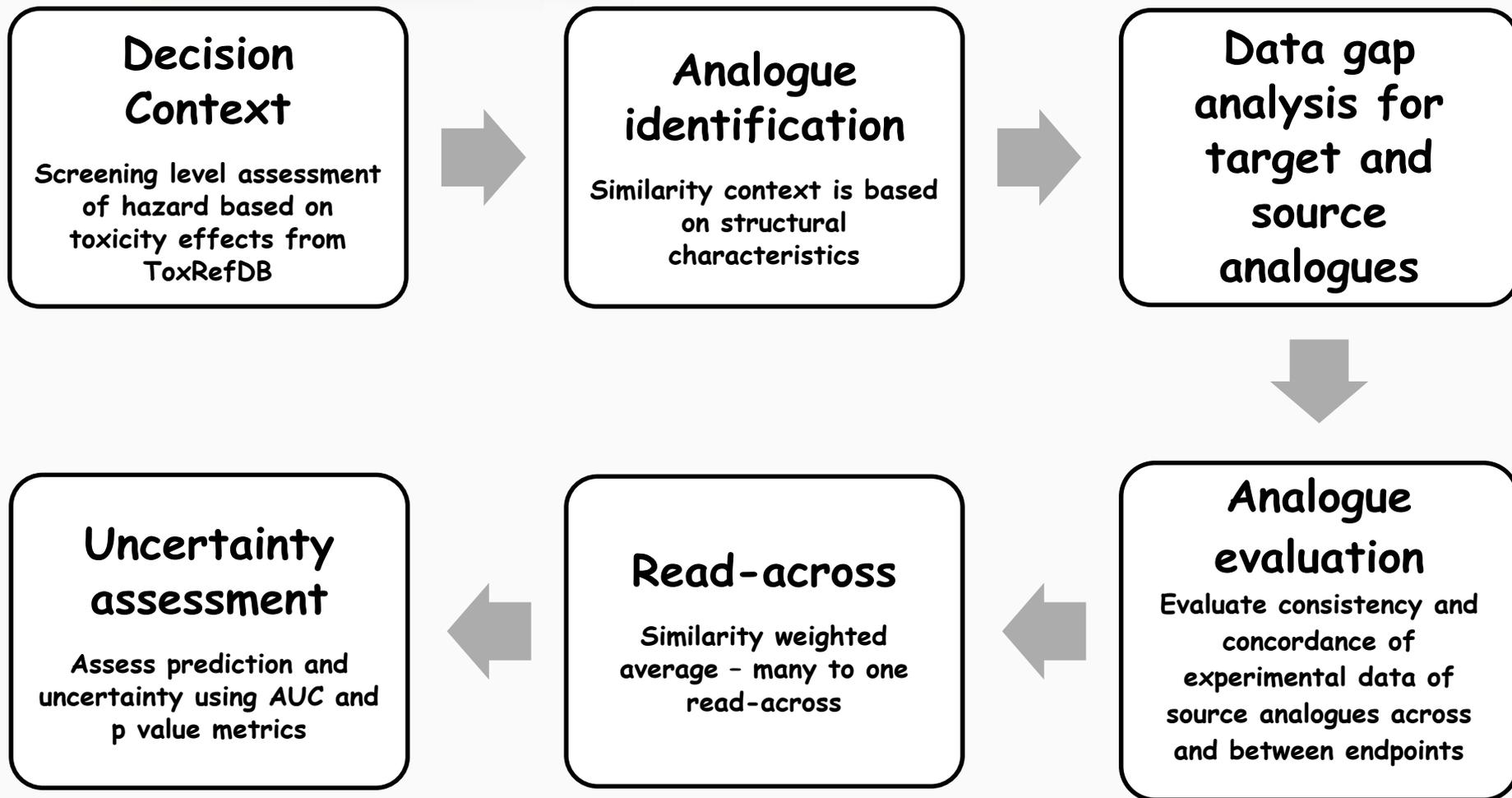
- Prediction of non-bioactivity
- Goal performance
- The for i



activity
'or
data
e

within specific study outcomes to be established

Read-across workflow in GenRA



GenRA (Generalised Read-Across)

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

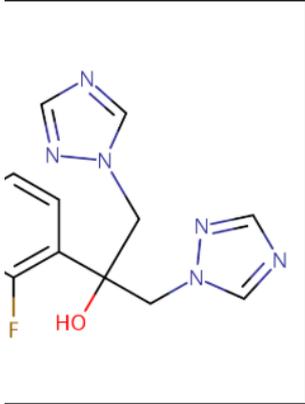
COMMENTS

Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

azole

-4 | DTXSID3020627
STox Substance Id.



Wikipedia

Fluconazole is an antifungal medication used for a number of fungal infections. This includes candidiasis, blastomycosis, coccidioidomycosis, cryptococcosis, histoplasmosis, dermatophytosis, and pityriasis versicolor. It is also used to prevent candidiasis in those who are at high risk such as following organ transplantation, low birth weight babies, and those with low blood neutrophil counts. It is given either by mouth or by injection into a vein. Common side effects include vomiting

...
[Read more](#)

Intrinsic Properties

Molecular Formula: $C_{12}H_{12}F_2N_6O$ [Mol File](#) [Find All Chemicals](#)

Average Mass: 306.277 g/mol [Isotope Mass Distribution](#)

Monoisotopic Mass: 306.104065 g/mol

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

GenRA (Generalised Read-Across)

Fluconazole

86386-73-4 | DTXSID3020627

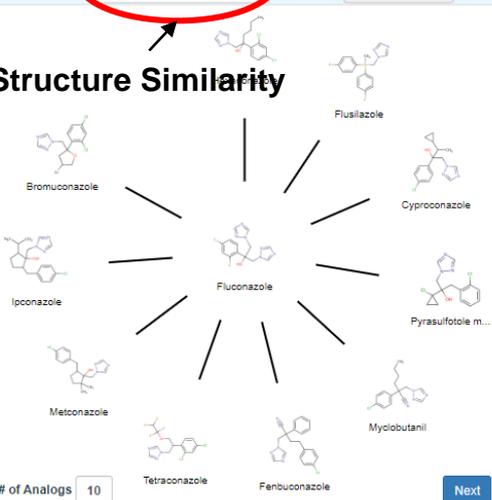
Searched by DSSTox Substance Id.

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

Step One: Analog Identification and Evaluation

Neighbors by: **Chem: Morgan Fgrprts** Filter by: invivo data

Structure Similarity



of Analogs: 10

Next

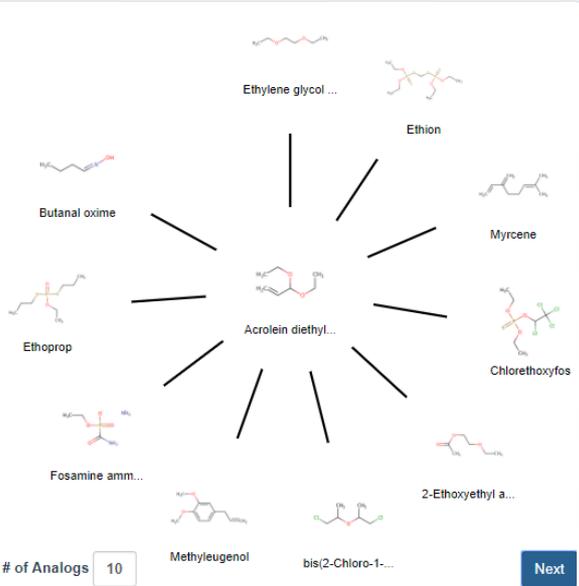
Select and Review Analogs

GenRA (Generalised Read-Across)

GenRA

Step Two: Data Gap Analysis & Generate Data Matrix

Neighbors by: Chem: Morgan Fgrpts Filter by: invivo data Summary Data Gap Analysis Group: ToxRef By: Tox Fingerprint **Generate Data Matrix**



	bio_k21	bio_kc1	chl_ct	tox_brf
Fluconazole	3	714	15	0
Hexaconazole	43	819	18	345
Flusilazole	28	819	9	345
Cyproconazole	14	819	16	408
Pyrasulfotole metabolite ...	0	0	18	234
Myclobutanil	15	818	15	345
Fenbuconazole	34	819	17	345
Tetraconazole	35	819	20	345
Metconazole	35	215	15	82
Ipcnazole	46	232	16	180
Bromuconazole	24	277	13	345

	Fluconazole	Hexaconazole	Flusilazole	Cyproconazole	Pyrasulfotole metab...	Myclobutanil	Fenbuconazole	Tetraconazole	Metconazole	Ipcnazole	Bromuconazole
CHR:Abdominal Cavity	█	█	█	█	█	█	█	█	█	█	█
CHR:Adrenal Gland	█	█	█	█	█	█	█	█	█	█	█
CHR:Artery (General)	█	█	█	█	█	█	█	█	█	█	█
CHR:Auditory Startle Re...	█	█	█	█	█	█	█	█	█	█	█
CHR:Bile duct	█	█	█	█	█	█	█	█	█	█	█
CHR:Blood	█	█	█	█	█	█	█	█	█	█	█
CHR:Blood vessel	█	█	█	█	█	█	█	█	█	█	█
CHR:Body Weight	█	█	█	█	█	█	█	█	█	█	█
CHR:Bone	█	█	█	█	█	█	█	█	█	█	█
CHR:Bone Marrow	█	█	█	█	█	█	█	█	█	█	█
CHR:Brain	█	█	█	█	█	█	█	█	█	█	█
CHR:Tracheus	█	█	█	█	█	█	█	█	█	█	█

Select and Review Analogs

Review Available Data

Fingerprint indicating available data

GenRA (Generalised Read-Across)

GenRA

Step Three: Run GenRA Prediction

Neighbors by: Chem: Morgan Fgrprts | Filter by: invivo data | Summary Data Gap Analysis | Group: ToxRef | By: Tox Fingerprint | Run Read-Across

Chemicals: Ethylene glycol..., Ethion, Butanal oxime, Myrcene

Summary Data Gap Analysis: Acrolein diethylacetal (14, 0, 4, 0), Ethylene glycol diethyl e... (7, 0, 4, 95)

Source analogues: Acrolein diethylacetal, Ethylene glycol diethyl e..., Ethion, Myrcene, Chloresthoxylfos, 2-Ethoxyethyl acetate, bis(2-Chloro-1-meth, Methylcyclohexal, Fosamine ammonium, Ethionop, Butanal oxime

Run Read-Across (circled in red)

Target

Similarity Weight: 0.39, 0.31, 0.29, 0.29, 0.26, 0.24, 0.22, 0.21, 0.21, 0.20 (circled in red)

Run GenRA

Organ	Fluconazole	Hexaconazole	Flusilazole	Cyproconazole	Pyrasulfotole m...	Myclobutanil	Fenbuconazole	Tetraconazole	Metconazole	Ipconazole	Bromuconazole
CHR:Abdominal Cavity	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
CHR:Adrenal Gland	Red	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Red
CHR:Artery (General)	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
CHR:Auditory Startle Re...	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
CHR:Bile duct	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
CHR:Blood	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
CHR:Blood vessel	Red	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
CHR:Body Weight	Red	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Red
CHR:Bone	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue

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

Demonstration



Related Publications

Chemical Research in Toxicology

Cite This: Chem. Res. Toxicol. 2017, 30, 2046-2059

Article

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



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

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

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. Lizzarraga^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

- 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

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

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>