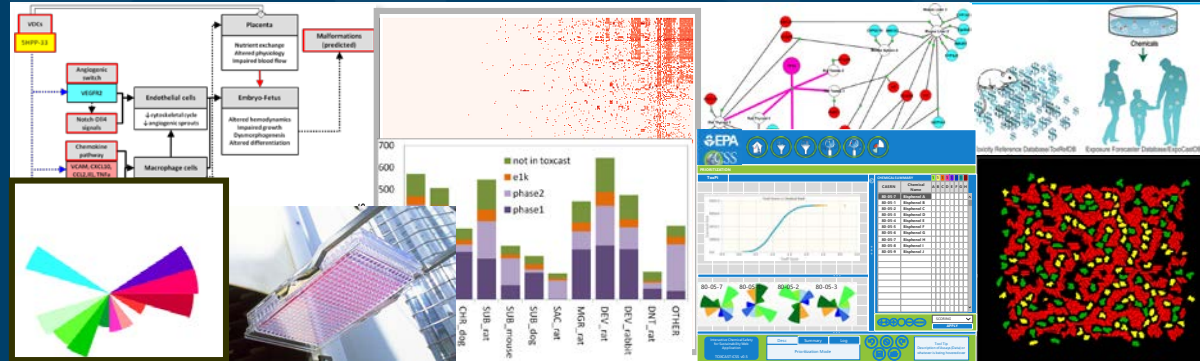


Quantifying uncertainty in read-across assessment – an algorithmic approach



SOT 2017 Workshop: Opportunities for read-across development and application using QSAR approaches
Tuesday 14th March 2017

Grace Patlewicz
National Center for Computational Toxicology

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

Conflict of Interest Statement

No conflict of interest declared.

Disclaimer:

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

Outline



- Background and Definitions
- Workflow for category development and read-across
- Current tools and approaches
- Uncertainty assessment in read-across
- Quantifying uncertainties and Assessing Performance of read-across
- From research to implementation
- Summary

Background & definitions

- Read-across describes one of the data gap filling techniques used within analogue and category approaches
- “Analogue approach” refers to grouping based on a very limited number of chemicals (e.g. target substance) + source substance)
- “Category approach” is used when grouping is based on a more extensive range of analogues (e.g. 3 or more members)

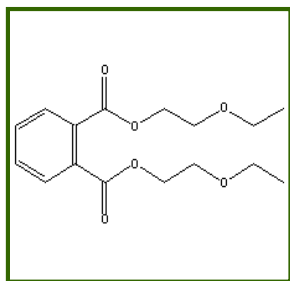
Definition: Read-across

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" i.e. endpoint & often study specific

	Source chemical	Target chemical
Property		

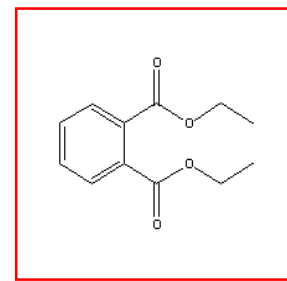
● Reliable data

○ Missing data



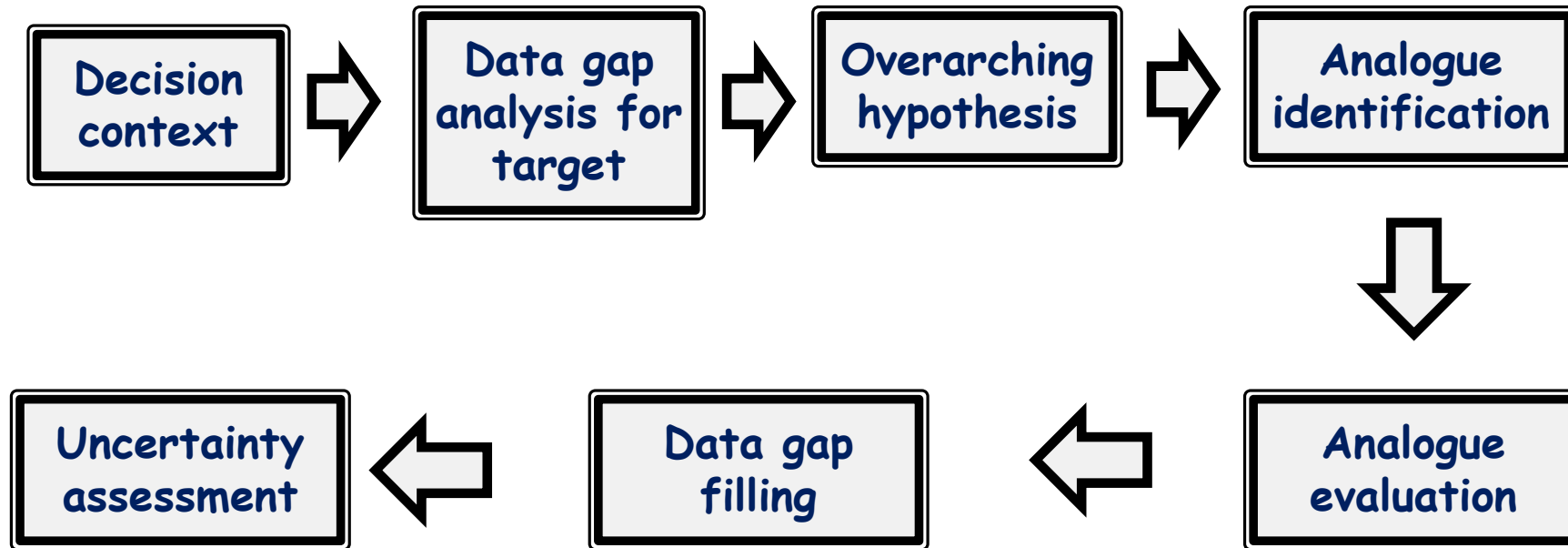
Known to be
harmful

Acute fish
toxicity?

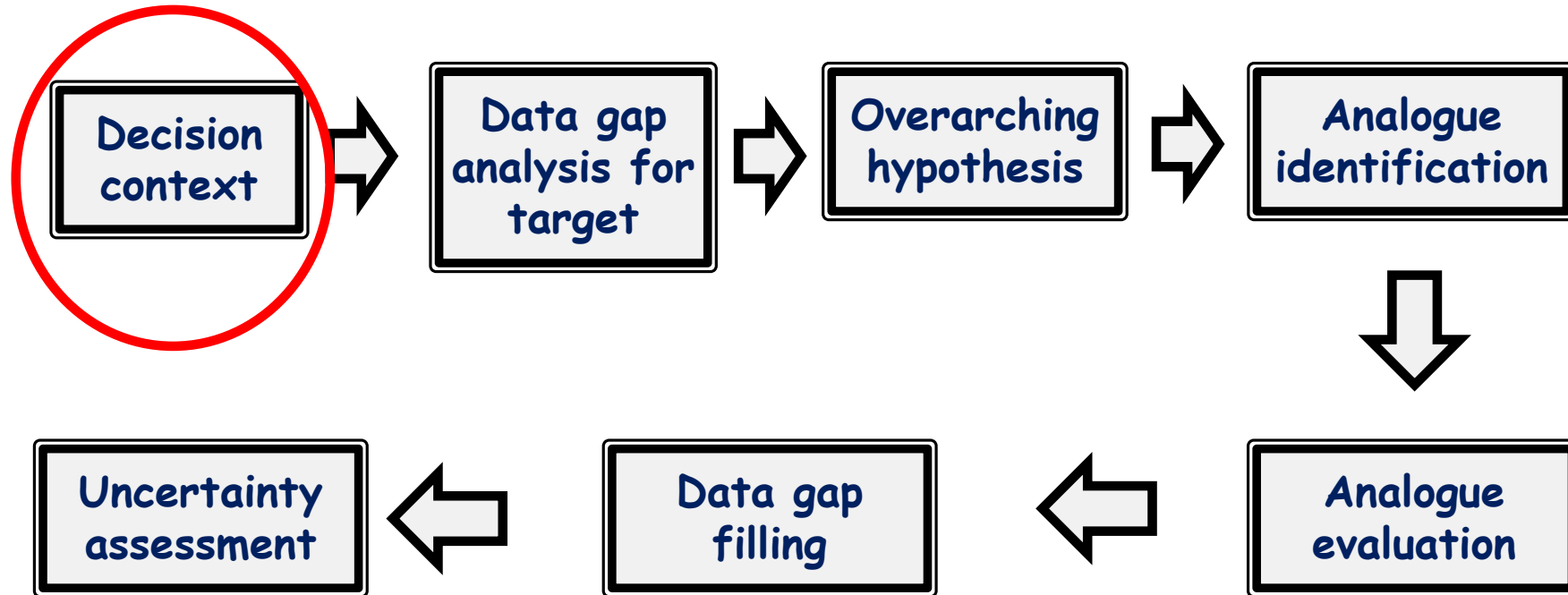


Predicted to be
harmful

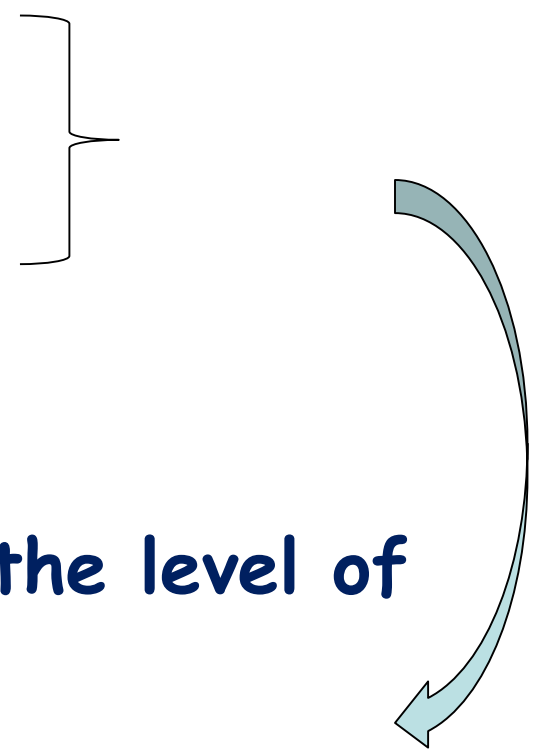
The Category Workflow



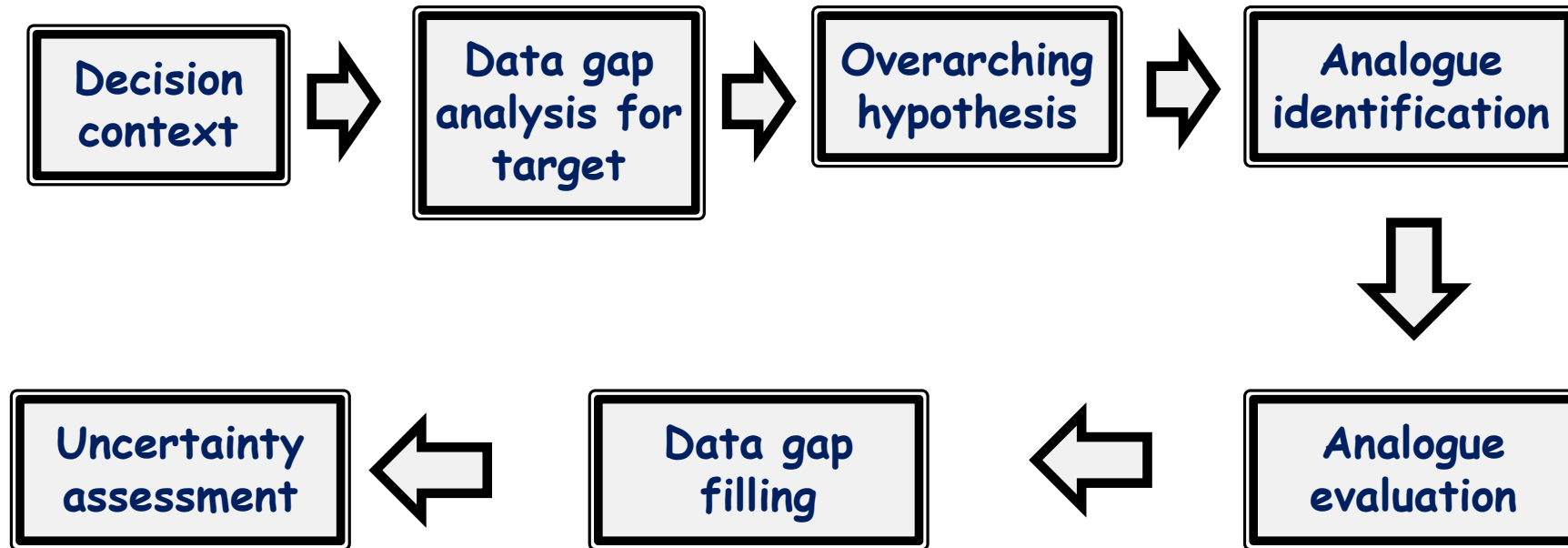
The Category Workflow



Decision Context

- Prioritisation, e.g. PMN
 - Screening level hazard assessment
 - Risk Assessment, e.g. PPRTV
- 
- Different decision contexts will dictate the level of uncertainty that can be tolerated

The Category Workflow



Selected Read-Across Tools

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

QSAR Toolbox 3.4.0.17 [Document]

Source substances

Target

Endpoint specific Similarity rationale

Data gap

Protein binding by OASIS
Protein binding by OECD
Protein binding potency
Supernatants
Toxic hazard classification by Cramer (ext)
Toxic hazard classification by Cramer (orig)
Ultimate biodeg
Biodeg BioHC half-life (Biowin)
Biodeg primary (Biowin 4)

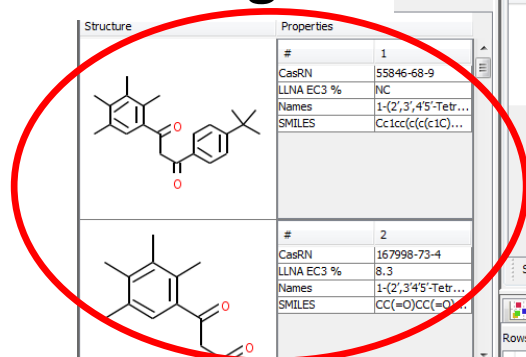
Defined Categories
Document
[481] AN2<AND>AN2 >> Michael addition to

Computational Toxicology

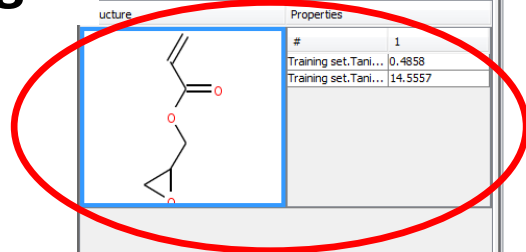
Structure	1	2	3	4	5	6	7	8	9
Immunotoxicity									
Irritation / Corrosion	(101/275)	M: not irritating, moderately irritating, n...	M: not irritating, no...	M: corrosive, corro...		M: irritating, corros...	M: slightly irritating...	M: moderately irrit...	
Neurotoxicity	(10/15)								
Photoinduced Toxicity									
Repeated Dose Toxicity	(69/6204)	M: 300 mg/kg bw/day (nominal), 0.5 mg/L	M: 15 mg/kg bw/d...	M: 10 mg/kg bw/d...		M: 55 mg/kg bw/d...	M: ≥124 mg/kg bw...	M: 20 mg/kg/day, ...	M: 3.33 mg/kg/c
Sensitisation									
Respiratory Tract	(1/1)								
Skin									
In Chemo									
In Vitro	(18/114)					M: 4.55 mg/L, 11.7...	M: <121 mg/L, <1...	M: sensitising, <4...	
In Vivo									
Alternative Methods	(1/1)								
Buehler Test	(5/5)						M: not sensitising		
Combined Intracutaneous and Topical S...	(1/1)								
Pattern	(1/1)								
e Adjuvant Test	(2/2)						M: NOT_SPECIFIED	M: not sensitising	
Lymph Node Assay	(12/14)						M: sensitising		
Sensitisation Test	(8/8)								
Lymph Node Assay	(1/1)						M: not sensitising	M: not sensitising	
Sensitisation Test	(46/64)		M: not sensitising...	M: sensitising		M: NOT_SPECIFIED	M: not sensitising...	M: sensitising	
Sensitisation Test	(4/6)					M: 4E3 µg/cm2, 1...		M: 400 µg/cm2, 1...	
Human Patch Test and Guinea Pig Mag...	(1/1)								
LLNA									
EC3	(20/31)					M: Positive	M: Positive	M: Negative	M: Positive
Maximization Test and Observations of ...	(1/1)								M: sensitising
Miscellaneous	(44/62)						M: Positive, Positiv...	M: Positive, Positive	
Modified Draize Test	(1/1)								
Modified Maximization Test	(1/1)								
Mouse Ear Swelling Test	(4/4)								
Mouse Local Lymphnode Assay (LLNA)									
Skin Sensitisation	(45/4)								
No Data	(1/1)						M: sensitising, NO...	M: sensitising	M: sensitising
Open Epicutaneous Test	(5/5)							M: not sensitising...	

Analogue identification & evaluation within Toxmatch

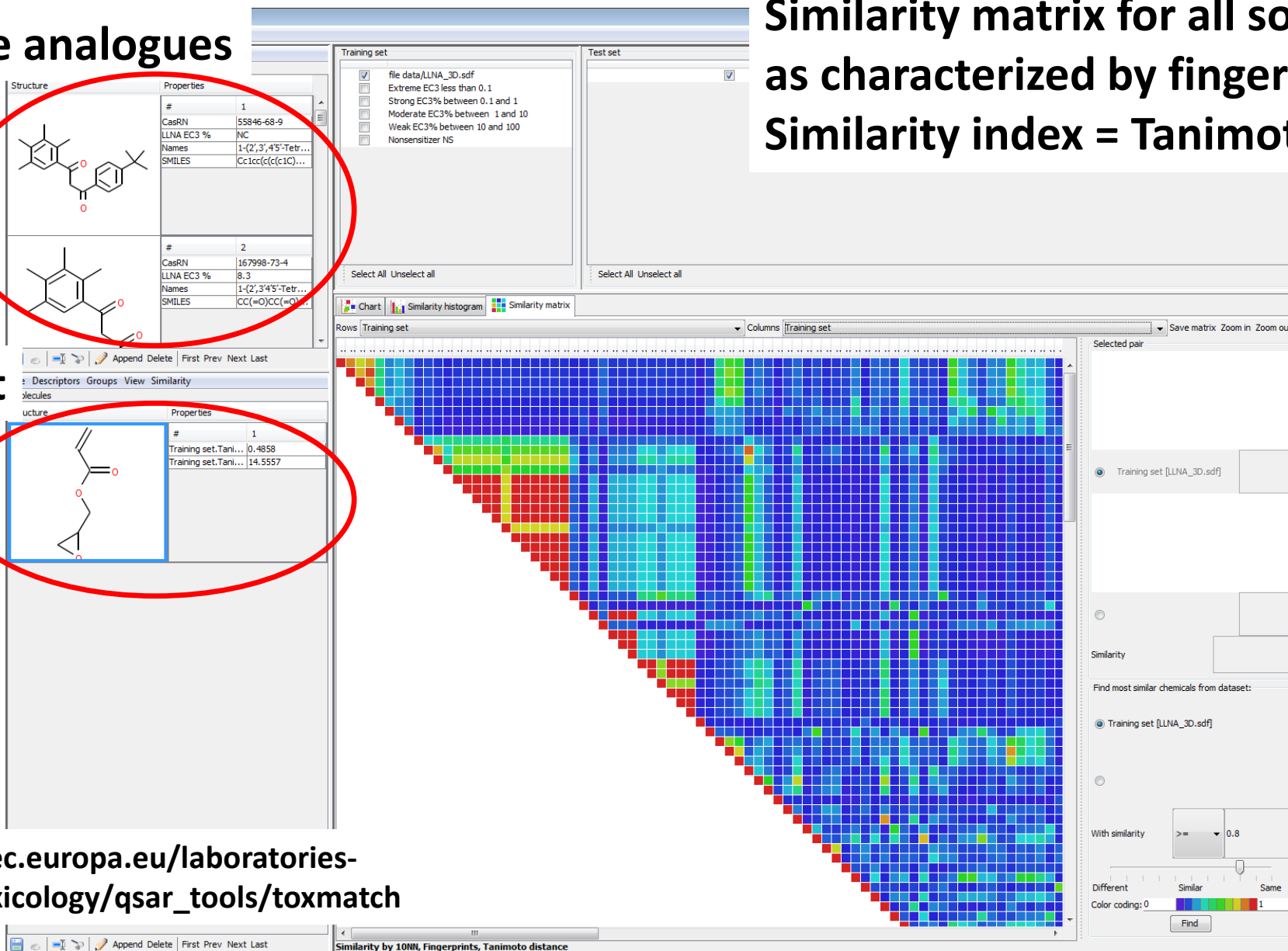
Source analogues



Target



Similarity matrix for all source analogues
as characterized by fingerprints
Similarity index = Tanimoto distance

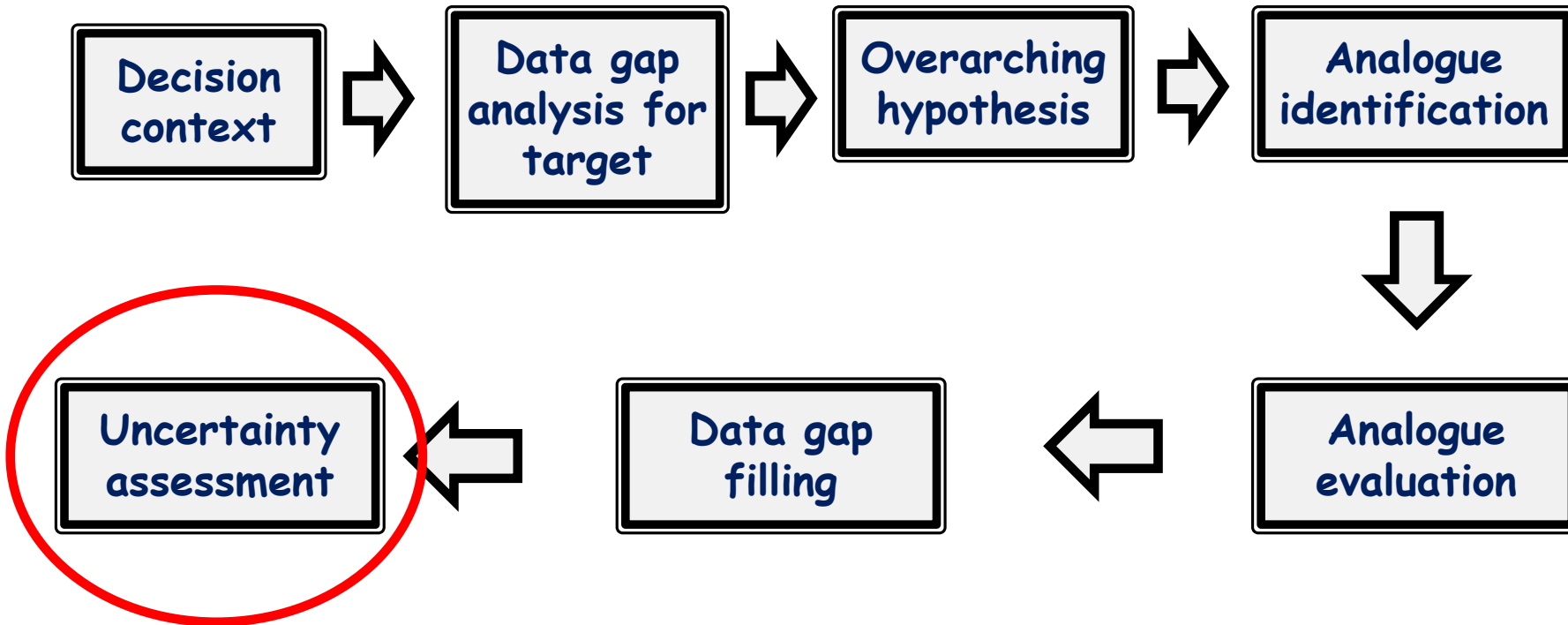


https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxmatch

Selected Read-Across Tools

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

The Category Workflow



Sources of Uncertainty

- Analogue or category approach? (# analogues)
- Completeness of the data matrix - no. of data gaps
- Data quality for the underlying analogues for the target and source analogues
- Consistency of data across the data matrix - concordance of effects and potency across analogues

Sources of Uncertainty (cont'd)

- Overarching hypothesis/similarity rationale - how to identify similar analogues and justify their similarity for the endpoint of interest
- Address the dissimilarities and whether these are significant from a toxicological standpoint e.g. ToxDelta
- Presence vs. absence of toxicity
- Toxicokinetics

Uncertainty Assessment

- A number of publications exist that can guide the construction and assessment of categories and use of read-across
 - Guidance and examples (OECD (2014), ECHA (2008), ECETOC (2012))
 - Frameworks for identifying analogues (e.g., Wu et al (2010), Patlewicz et al (2013))
 - Frameworks for assessing read-across (Blackburn and Stuard (2014), Patlewicz et al (2014), Patlewicz et al (2015), ECHA - RAAF (2015), Schultz et al (2015), Ball et al (2016))

Uncertainty assessment

- However read-across acceptance relies on a subjective expert assessment
- There is no objective measure of read-across performance
- Different approaches have been explored to characterise uncertainties both qualitatively and quantitatively
- E.g. Blackburn and Stuard, Molecular Networks, EPA NCCT

Uncertainty assessment

- **Low** - degree of uncertainty is judged to be comparable to having direct data on a chemical

Regulatory Toxicology and Pharmacology 68 (2014) 353–362



Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph



A framework to facilitate consistent characterization of read across uncertainty



Karen Blackburn*, Sharon B. Stuard|

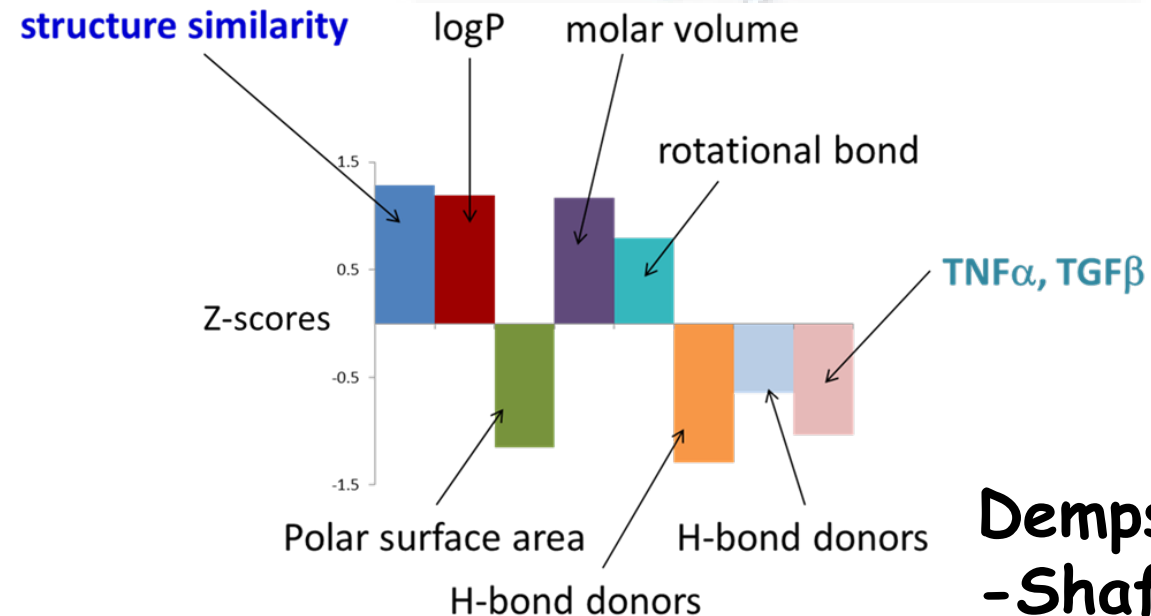
Central Product Safety Department, The Procter & Gamble Company, Mason Business Center, 8700 Mason Montgomery Road, Cincinnati, OH 45040, United States

- **High** - degree of uncertainty is judged to be significantly greater than if direct data were used
 - Read across is not actionable without additional information

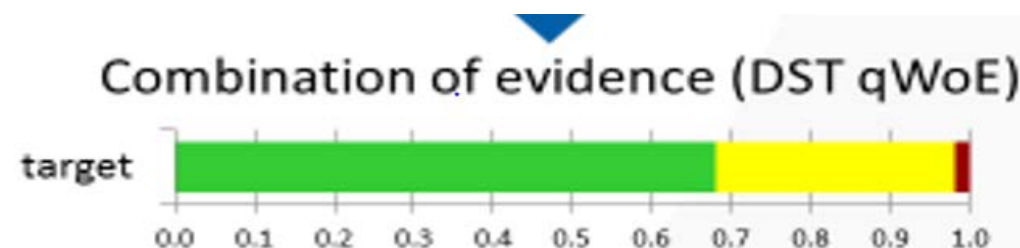
Setting up the table of reliability metrics

INFORMATION TYPE: Analogs, In vivo toxicity data, Knowledgebase

Target/Analog		
Experimental	data	reproductive developmental
	reliability	Klimisch score
Analog similarity:		structure properties biological assays
Analog similarity:		Skyline profile
Predictions alerts/QSAR): endpoints		reproductive developmental
Reliability		PPV/NPV Odd ratio



Dempster-Shafer Theory



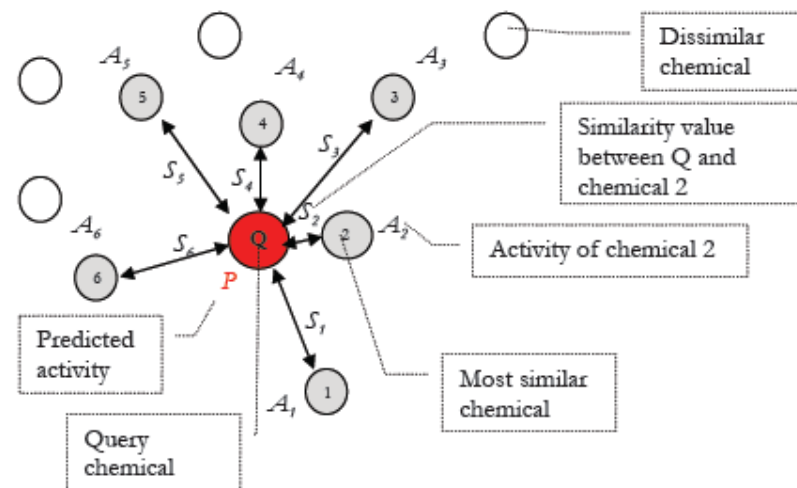
Quantifying Uncertainty & Assessing Performance of Read-Across

- GenRA (Generalised Read-Across) is a “local validity” approach
- Predicting toxicity as a similarity-weighted activity of nearest neighbors based on chemistry and bioactivity descriptors
- Systematically evaluates read-across performance and uncertainty using available data

$$y_i^{\beta, \alpha} = \frac{\sum_j^k s_{ij}^{\alpha} x_j^{\beta}}{\sum_j^k s_{ij}^{\alpha}}$$

Jaccard similarity:

$$s_{ij} = \frac{\sum_l (x_{il} \wedge x_{jl})}{\sum_l (x_{il} \vee x_{jl})}$$



GenRA - Approach

I. Data

1,778 Chemicals
3,239 Structure descriptors (chm)
820 Bioactivity assays (bio)
ToxCast
574 Apical outcomes (tox)
ToxRefDB

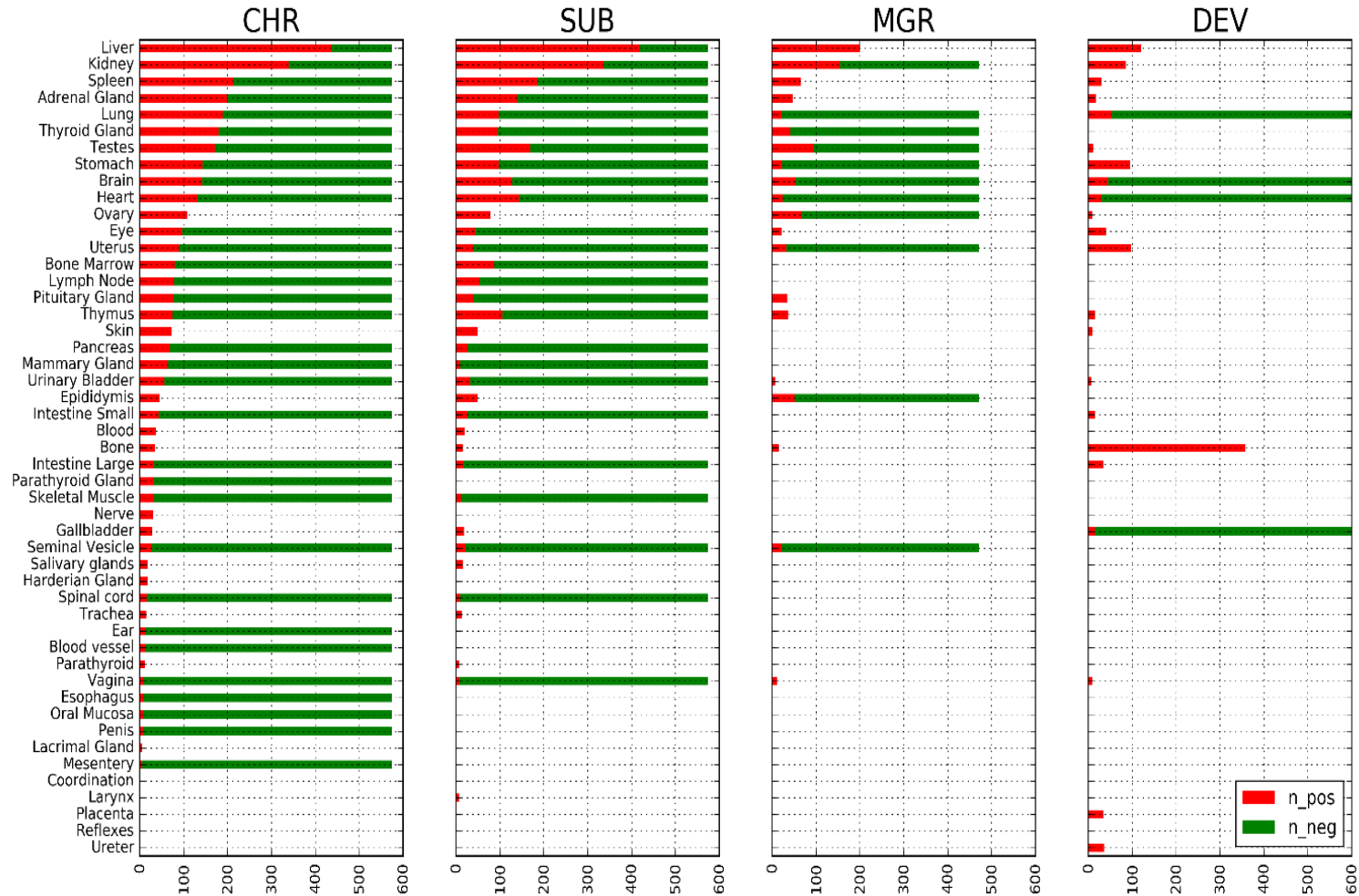
II. Define Local neighborhoods

Use K-means analysis to group chemicals by similarity
Use cluster stability analysis
~ 100 local neighborhoods

III. GenRA

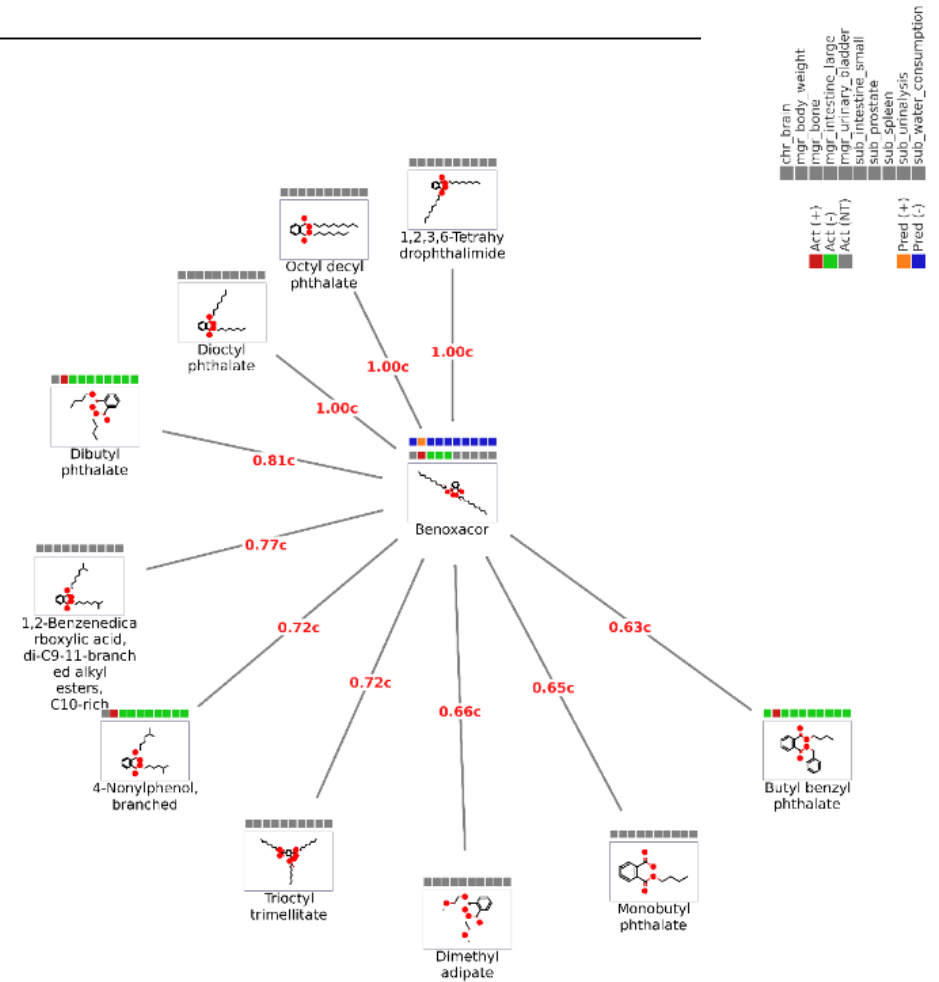
Use GenRA to predict apical outcomes in local neighborhoods
Evaluate impact descriptors (chm, bio, bc) on prediction
Quantify uncertainty

GenRA - Toxicity Data from ToxRefDB

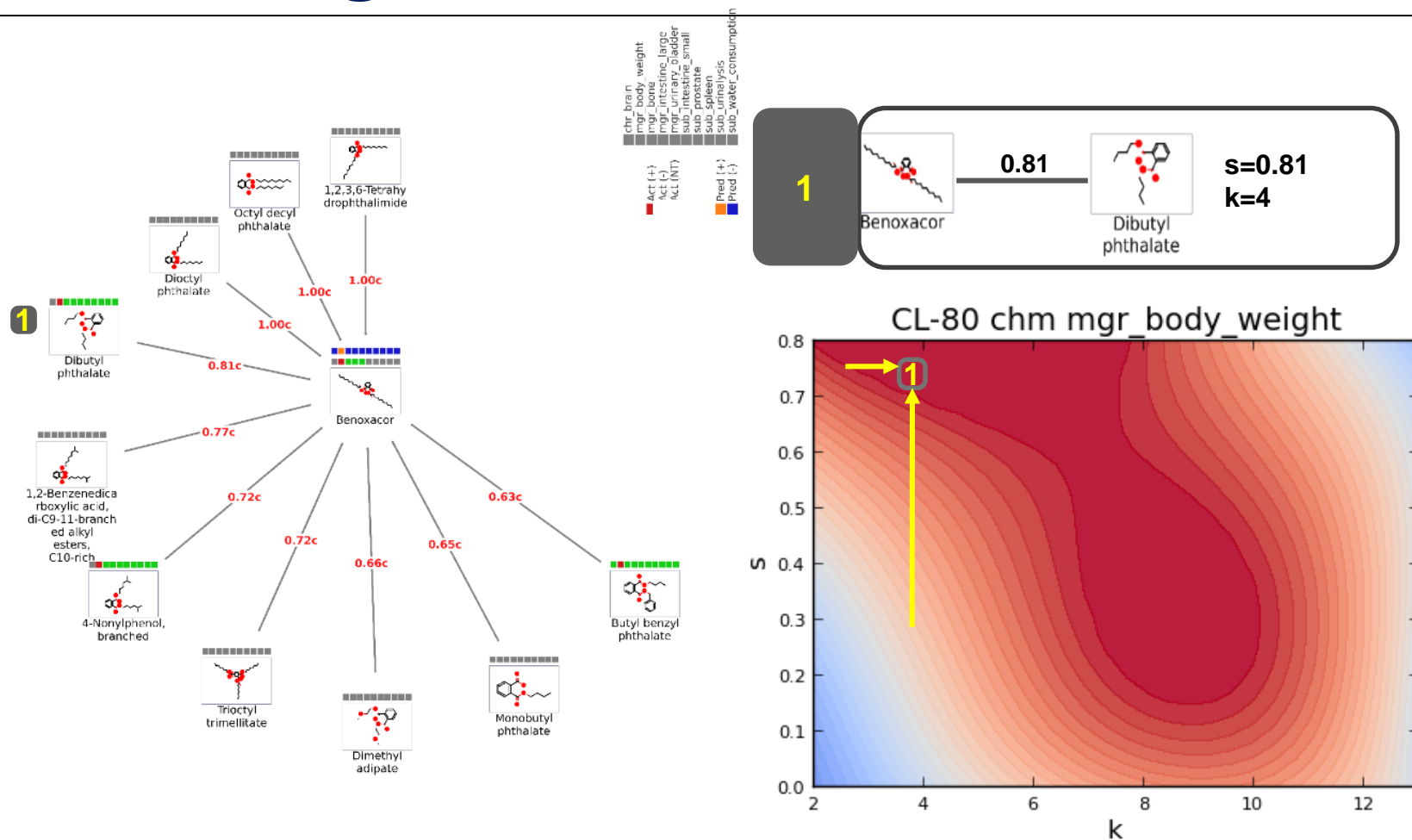


GenRA – Performance in Each Cluster

- Use GenRA to predict the similarity weighted toxicity scores for each
 - Toxicity type (β)
 - Descriptor = {chm, bio, bc} (α)
 - No. of nearest neighbors (k)
 - Similarity score threshold (s_{ij}^{α})
- Calculate performance by comparing predicted y^{tox} and true x^{tox} for all chemicals using area under ROC curve (AUC)
- Results: {cluster, α , β , k , s , AUC}



GenRA - Analysing Local Neighborhood of a Chemical



GenRA – Insights and Next Steps

- The approach enabled a performance baseline for read-across predictions of specific study outcomes to be established but was still context dependent on the endpoint and the chemical
- Ongoing analysis:
- Consideration of other information to refine the analogue selection – e.g. TK similarity, metabolic similarity, reactivity similarity...

- Inter
ongoi

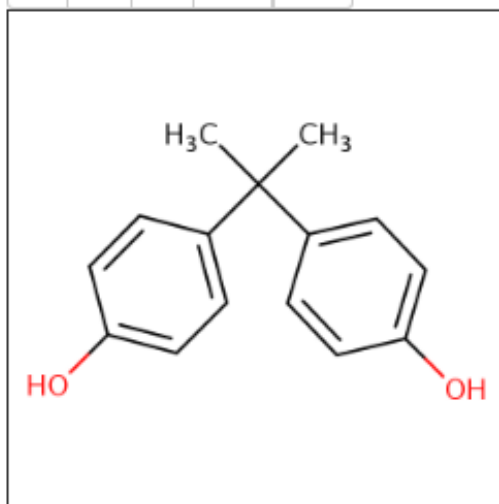
of

Basic Integration via GenRA tab

Bisphenol A

80-05-7 | DTXSID7020182

🔍 Searched by Approved Name: Found 1 result for 'bisphenol A'.



Insert GenRA Tab and
Appropriate Sub-tabs



Wikipedia

Bisphenol A (BPA) is an organic synthetic compound with the chemical formula Cc1ccc(cc1)C(C)(C)c2ccc(O)cc2. It is soluble in organic solvents, but poorly soluble in water. It has been in common use since the 1960s. BPA is employed to make certain plastics and epoxy resins. BPA-based plastics are used in a wide range of consumer products, including polycarbonate plastic bottles, food storage containers, and thermal paper receipts.

Intrinsic Properties

Structural Identifiers

Record Information

Chemical Properties

External Links

Synonyms

Env. Fate/Transport

Bioassays

Exposure

Analytical

Literature

Similar Molecules

Comments

Summary

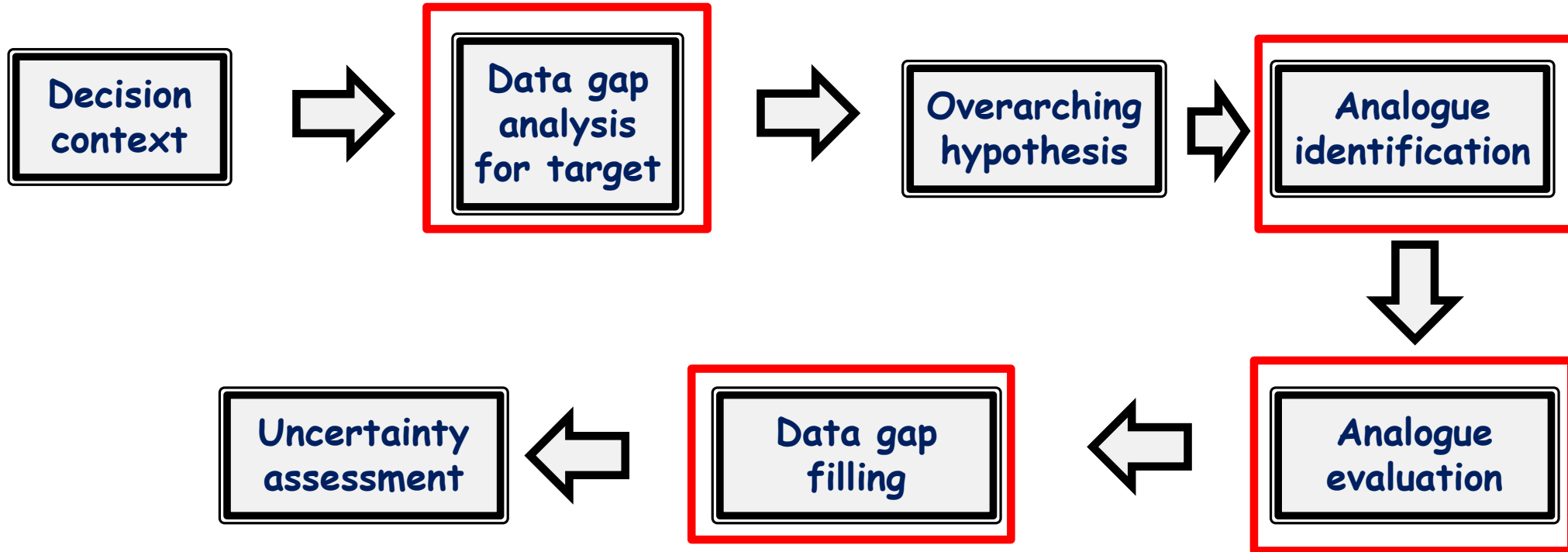
Download as:

TSV

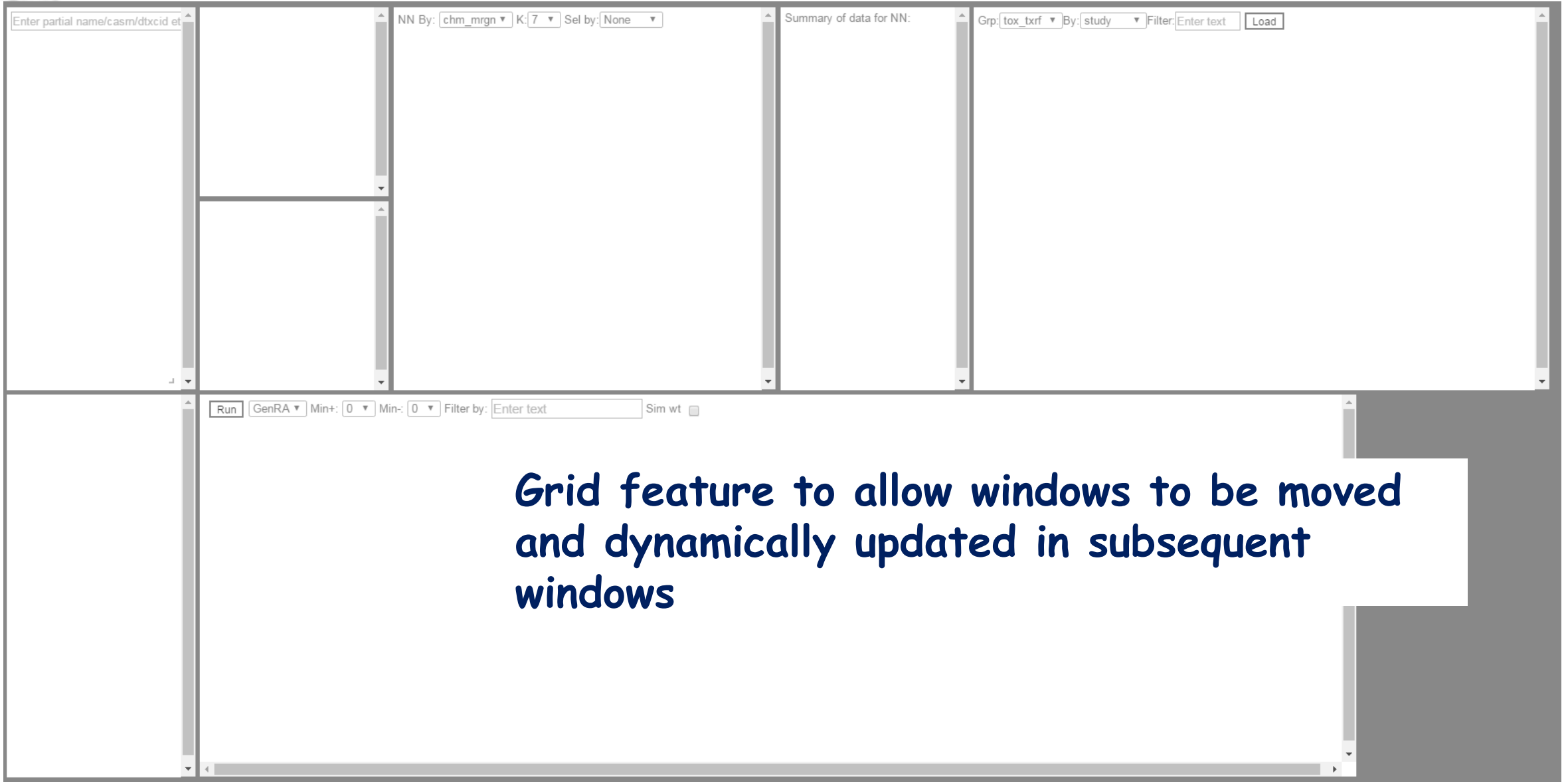
Excel

SDF

GenRA prototype development



Initial interface



The interface is a web-based application with a grid layout. It features several panels for data entry and analysis. The top row contains four panels: a search panel on the left, a panel for NN (Neural Network) settings, a summary panel, and a filter panel. The bottom row contains a large panel for results and a panel for GenRA settings. The interface is designed to be flexible, allowing users to move and resize panels.

Enter partial name/casm/dxcid et

NN By: chm_mrgn K: 7 Sel by: None


Summary of data for NN:

Grp: tox_txrf By: study Filter: Enter text Load

Run GenRA Min+: 0 Min-: 0 Filter by: Enter text Sim wt


Grid feature to allow windows to be moved and dynamically updated in subsequent windows

1569-01-3



1-Propoxy-2-propanol
CASRN: 1569-01-3
ID: DTXCID409217

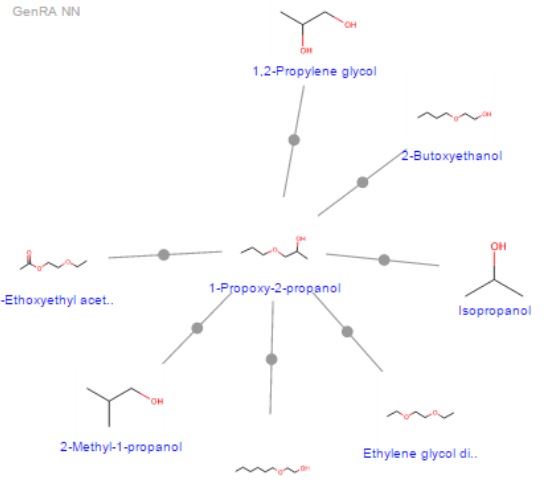
1-Propoxy-2-propanol | 1569-01-3



dsstox_cid:DTXCID409217
dsstox_sid:DTXSID5029217
gsid:29217
inchi_key:FENFUOGYJVOCRY-UHFFFAO
lupac:1-propoxypropan-2-ol
mol_weight:118.17418381835938
pubchem_cid:15286
smiles:CCOCC(C)O
tag:TOXCST,CTD

NN By: chm_mrgn K: 7 Sel by: tox_txf

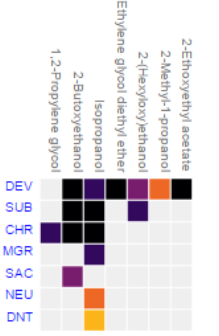
GenRA NN



Summary of data for NN:

	tox_txf	chm_at	bio_txet	bio_tx21
Ethylene glycol diet				
1-Propoxy-2-propanol				
Isopropanol				
2-Methyl-1-propanol				
2-(Hexyloxy)ethanol				
2-Ethoxyethyl acetat				
1,2-Propylene glycol				
2-Butoxyethanol				

Grp: tox_txf By: study Filter:



Run GenRA Min+: 0 Min-: 0 Filter by: Sim wt ☐

1569-01-3

1-Propoxy-2-propanol | 1569-01-3

1-Propoxy-2-propanol
CASRN: 1569-01-3
ID: DTXCID409217

Chemical structure of 1-Propoxy-2-propanol: CCOC(C)CO

dsstox_cid:DTXCID409217
dsstox_sid:DTXSID5029217
gsid:29217
inchi_key:FENFUOGYJVOCRY-UHFFFAO
iupac:1-propoxypropan-2-ol
mol_weight:118.17416381835938
pubchem_id:15288
smiles:CCOCC(C)O
tag:TOXCST,CTD

NN By: chm_mrgn K: 7 Sel by: None

Summary of data for NN:

Grp: tox_txrf By: study Filter: Enter text Load

Sim wt ☐

Introduce and select a target chemical

1569-01-3

CCOC(C)CO
1-Propoxy-2-propanol
CASRN: 1569-01-3
ID: DTXCID409217

dsstox_cid:DTXCID409217
dstox_sid:DTXSID5029217
sid:29217
inchi_key:FENFUOGYJVOCRY-UHFFFAO
iupac:1-propoxypropan-2-ol
mol_weight:118.17416381835938
pubchem_cid:15286
smiles:CCOCC(C)O
tag:TOXCST,CTD

1-Propoxy-2-propanol | 1569-01-3

GenRA NN

Similarity index

Summary of data for NN:

	tox_txrf	chm_ct	bio_txrf	bio_txct
Ethylene glycol diet				
1-Propoxy-2-propanol				
Isopropanol				
2-Methyl-1-propanol				
2-(Hexyloxy)ethanol				
2-Ethoxyethyl acetat				
1,2-Propylene glycol				
2-Butoxyethanol				

Grp: tox_txrf By: study Filter: Enter text Load

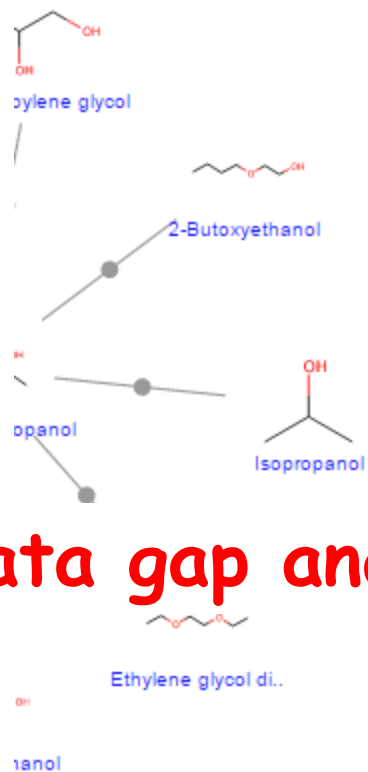
	2-Ethoxyethyl acetate	2-Methyl-1-propanol	2-(Hexyloxy)ethanol	Ethylene glycol diethyl ether	Isopropanol	2-Butoxyethanol	1,2-Propylene glycol
DEV							
SUB							
CHR							
MGR							
SAC							
NEU							
DNT							

Run GenRA Min+: 0 Min-: 0 Filter by: Enter text Sim wt

Analogue identification:

Search for source analogues
on the basis of chemical
fingerprints, filtered by
availability of in vivo data

Grp: tox_txrf



Summary of data for NN:

	bio_tx21	bio_txct	chm_ct	tox_txrf
Ethylene glycol diet				
1-Propoxy-2-propanol				
Isopropanol				
2-Methyl-1-propanol				
2-(Hexyloxy)ethanol				
2-Ethoxyethyl acetat				
1,2-Propylene glycol				
2-Butoxyethanol				

Grp: tox_txrf By: study Filter: Enter text Load

	1,2-Propylene glycol	2-Butoxyethanol	Isopropanol	Ethylene glycol diethyl ether	2-(Hexyloxy)ethanol	2-Methyl-1-propanol	2-Ethoxyethyl acetate
DEV							
SUB							
CHR							
MGR							
SAC							
NEU							
DNT							

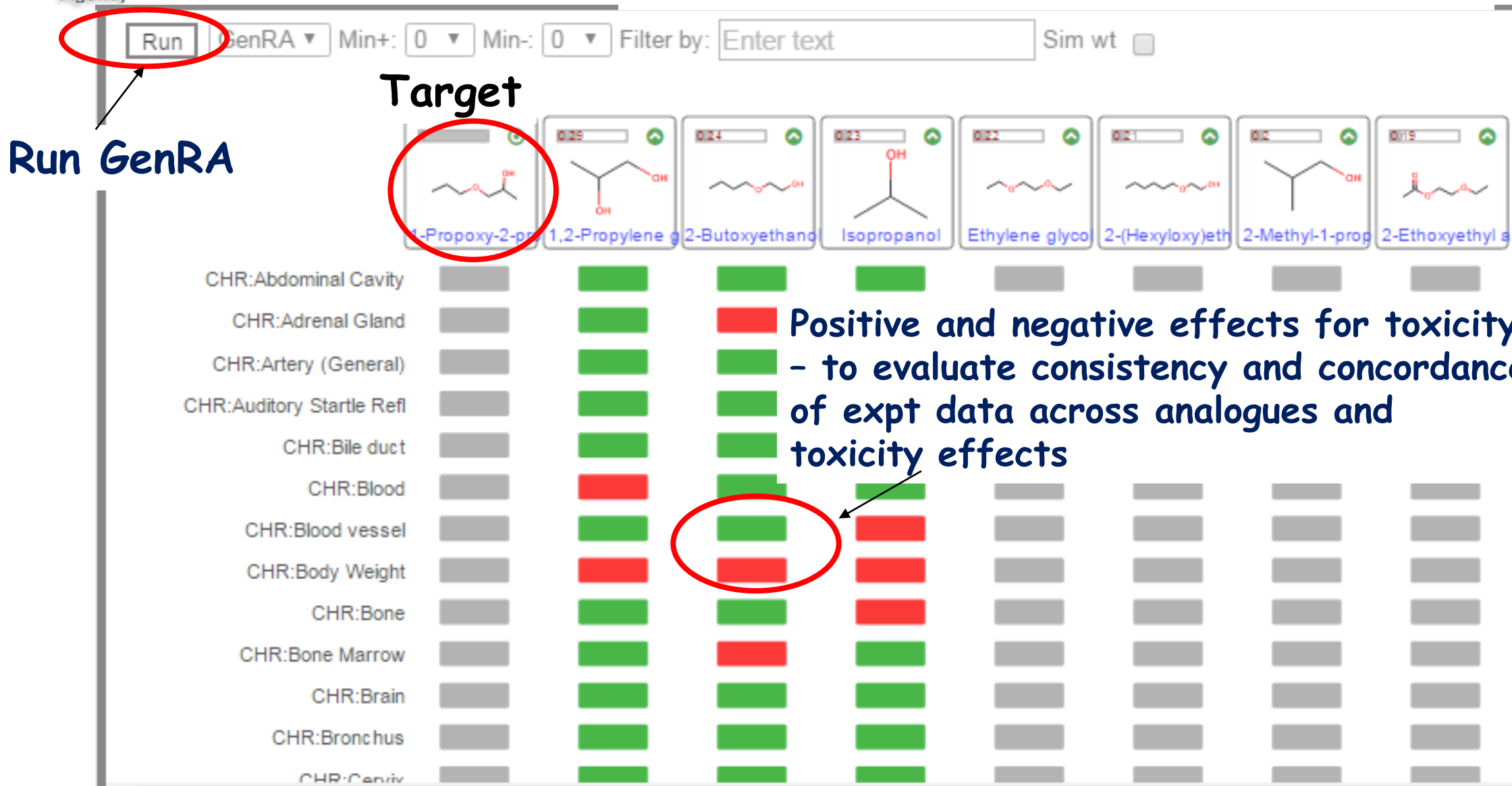
To initiate data matrix view

Data gap analysis

Data gap analysis - View data coverage across study type on the basis of toxicity effects

View data quantity by type

Analogue evaluation using data matrix view



Data gap filling using GenRA within data matrix



Summary

- Still many challenges remain in read-across – what information is relevant to integrate and ways in which that integration can be performed
- Quantifying the uncertainty of read-across prediction is a critical issue
- Have illustrated the research directions being taken within NCCT and work to implement these into practical tools
- To see more – stop by the EPA booth in the ToxExpo for live demonstrations of the CompTox dashboard, GenRA (Wed 1pm) and more..

Acknowledgements

- Imran Shah
- George Helman
- Tony Williams
- Jeff Edwards
- Richard Judson
- Chris Grulke
- Ann Richard
- Karen Blackburn P&G
- Sharon Stuard P&G
- Chihae Yang Altamira LLC & Molecular Networks