The use of in silico tools for weight of evidence

Emilio Benfenati

ISTITUTO DI RICERCHE
FARMACOLOGICHE
MARIO NEGRI · IRCCS

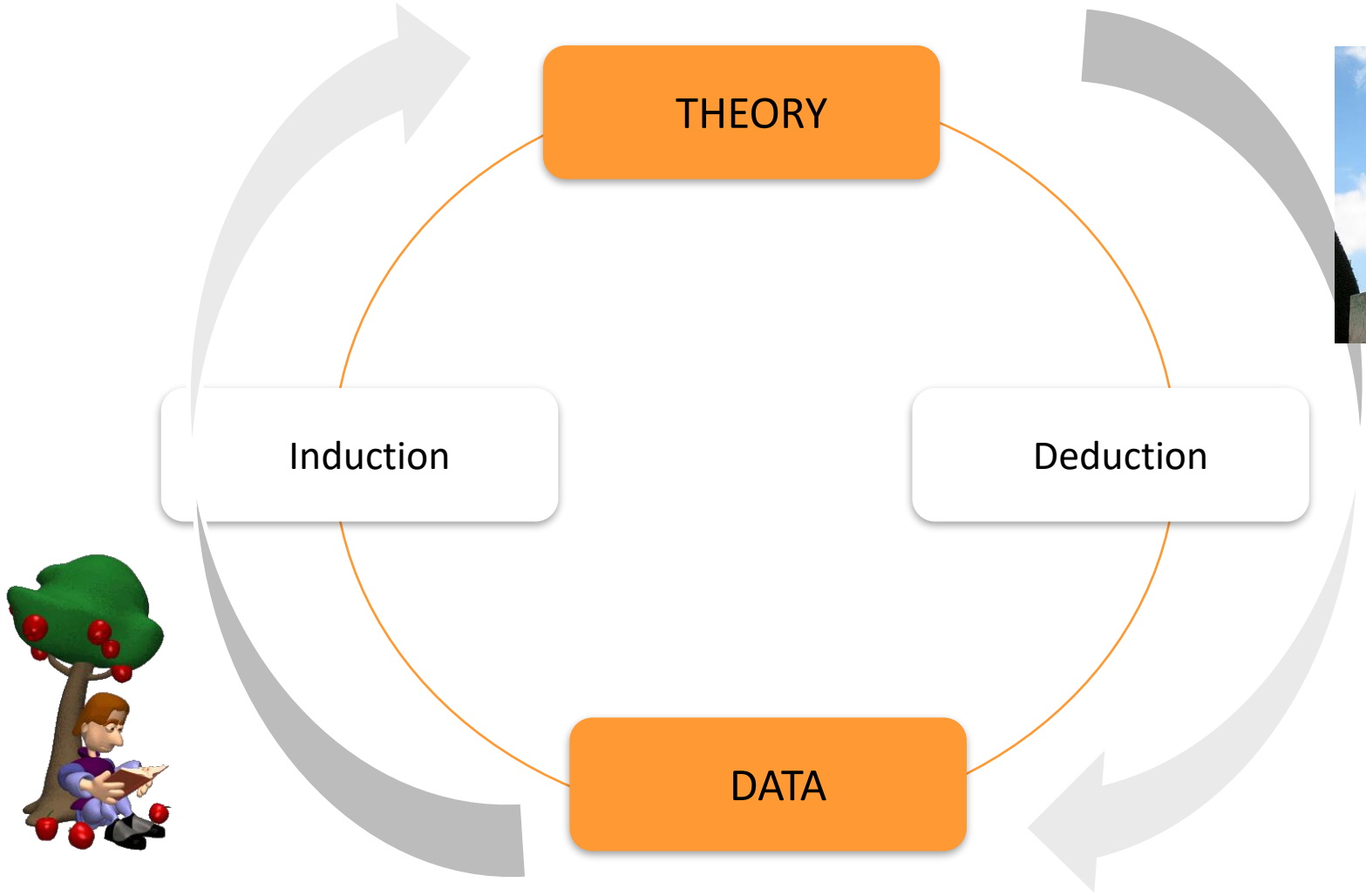
The gap between authorities and in silico models

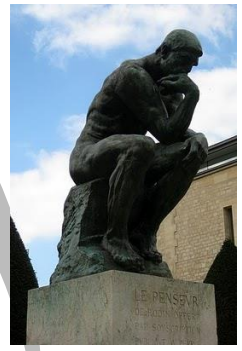
Authorities use:

- data
- theoretical explanation

In silico models provide: predictions

Integrating observation and theory (ii)





The perspective: to provide data and theory in an integrated way

- Heterogeneous inputs
- Integration?

Weight of evidence (WoE): EFSA Guidance

SCIENTIFIC OPINION

EFSA Journal

ADOPTED: 12 July 2017

doi: 10.2903/j.efsa.2017.4971

Guidance on the use of the weight of evidence approach in scientific assessments

EFSA Scientific Committee,
Anthony Hardy, Diane Benford, Thorhallur Halldorsson, Michael John Jeger,
Helle Katrine Knutsen, Simon More, Hanspeter Naegeli, Hubert Noteborn, Colin Ockleford,
Antonia Ricci, Guido Rychen, Josef R Schlatter, Vittorio Silano, Roland Solecki,
Dominique Turck, Emilio Benfenati, Qasim Mohammad Chaudhry, Peter Craig,
Geoff Frampton, Matthias Greiner, Andrew Hart, Christer Hogstrand, Claude Lambre,
Robert Luttik, David Makowski, Alfonso Siani, Helene Wahlstroem, Jaime Aguilera,
Jean-Lou Dorne, Antonio Fernandez Dumont, Michaela Hempen, Silvia Valtueña Martínez,
Laura Martino, Camilla Smeraldi, Andrea Terron, Nikolaos Georgiadis and Maged Younes

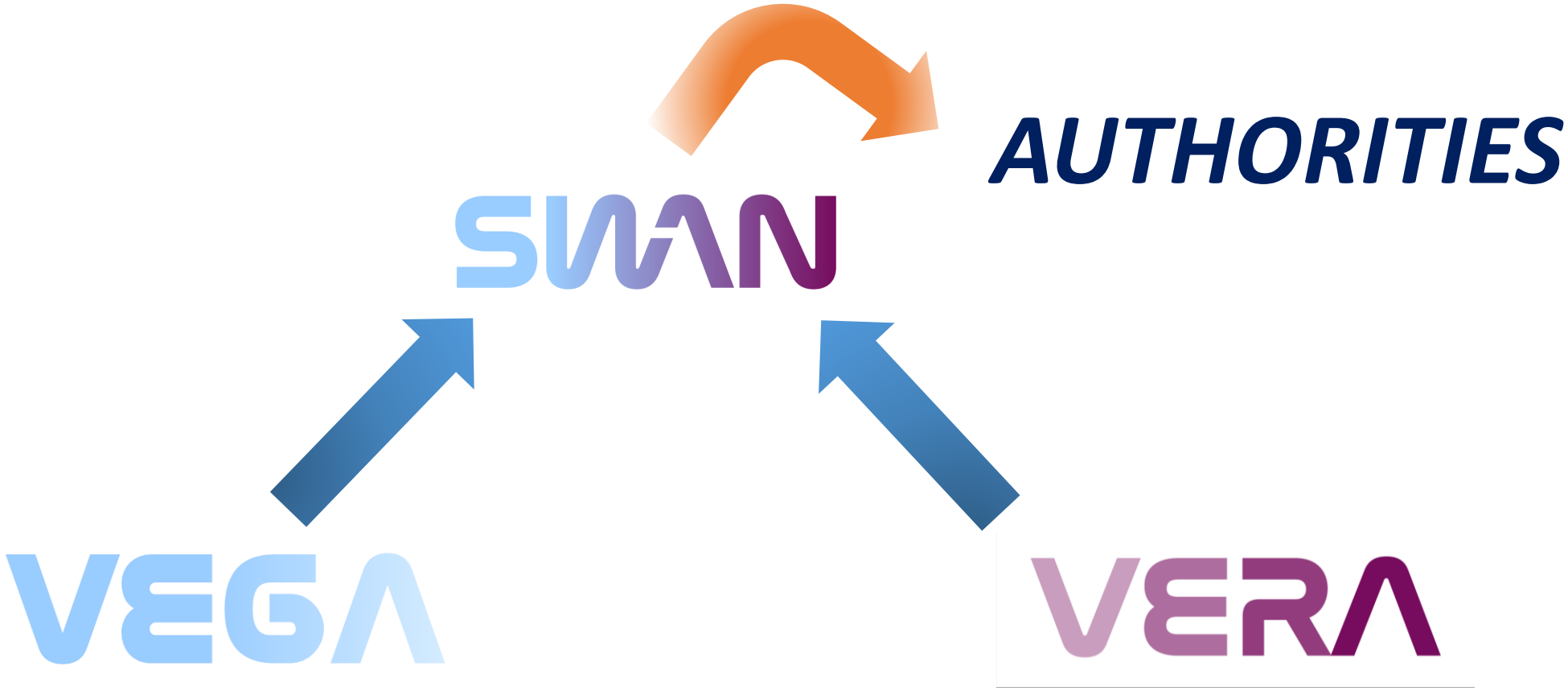
https://www.efsa.europa.eu/en/efsajournal/pub/4971

The TWO predictions

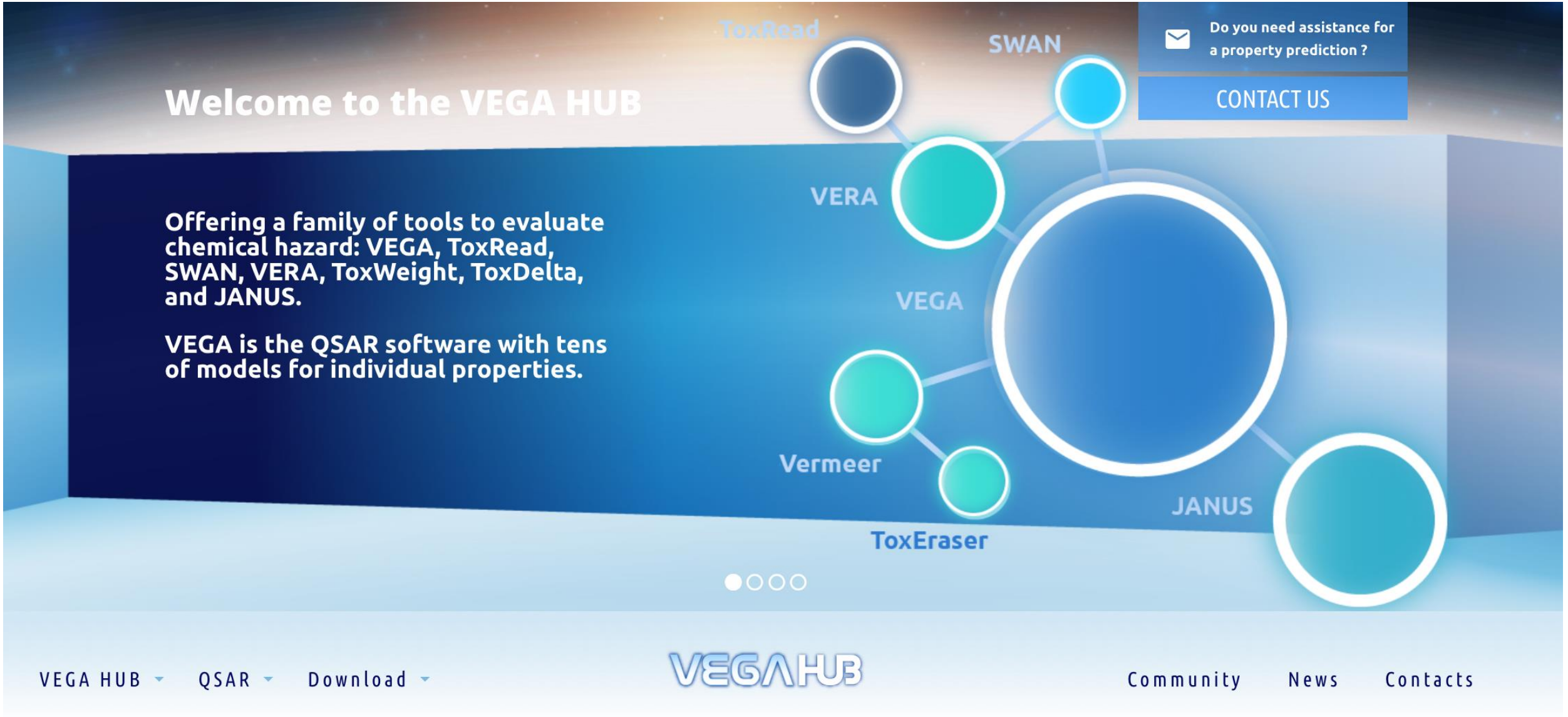
- To predict the **property value** of the substance - unknown value, I may know the (potential) mechanism

- To predict the **mechanism** – I may have all tox values of all substances, but ignore the mechanism.

SWAN, the door between the two worlds: real and virtual

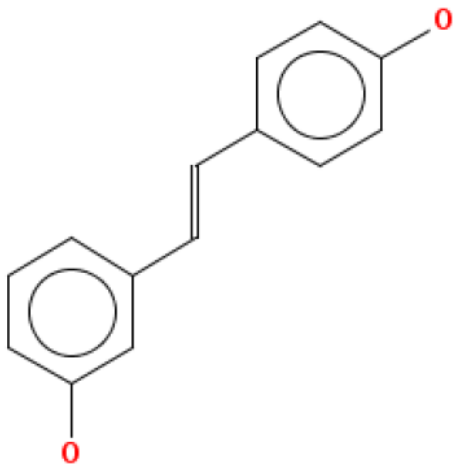


VEGAHUB – the different tools



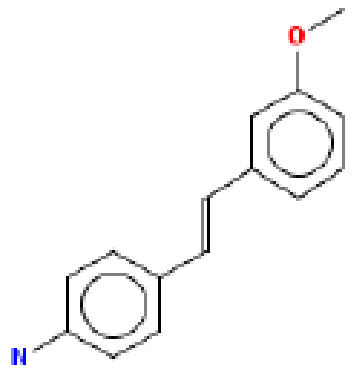
https://www.vegahub.eu/

VEGA – the info on data and mechanism



Prediction: Reliability:

Prediction is Possible NON-Mutagen
critical aspects, which require to be
- similar molecules found in the trair
that disagree with the predicted valu

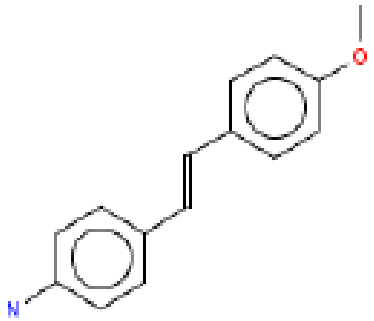


Compound #1

CAS: 154028-32-7
Dataset id: 2989 (Training set)
SMILES: O(c2cccc(C=Cc1ccc(N)cc1)c2)C
Similarity: 0.907

Experimental value: Mutagenic
Predicted value: Mutagenic

Alerts (not found in the target): SM44; SM104



Compound #2

CAS: 7570-37-8
Dataset id: 1345 (Training set)
SMILES: O(c1ccc(cc1)C=Cc2ccc(N)cc2)C
Similarity: 0.905

Experimental value: Mutagenic
Predicted value: Mutagenic

Alerts (not found in the target): SM44; SM104

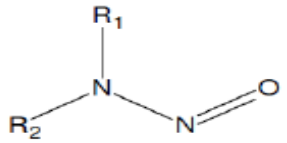
Compound #3

VEGA – the mechanism and the data

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 3) Reasoning on fragments/structural alerts:

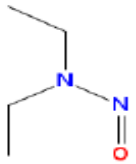
Fragment found: SA21 Alkyl and aryl N-nitroso groups



R1= Aliphatic or aromatic carbon,
R2= Any atom/group

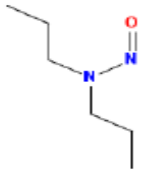
Alkyl and aryl N-nitroso groups

Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 55-18-5
Dataset id: 516 (Training set)
SMILES: O=NN(CC)CC
Similarity: 1

Experimental value: Mutagenic
Predicted value: Mutagenic



CAS: 621-64-7
Dataset id: 520 (Training set)
SMILES: O=NN(CCC)CCC
Similarity: 0.906

Experimental value: Mutagenic
Predicted value: Mutagenic

CAS: 38434-77-4

The reliability of the predictions. ADI

3.2 Applicability Domain:
Measured Applicability Domain Scores

Global AD Index
AD index = 0.979
Explanation: the predicted compound is into the Applicability Domain of the model.

Similar molecules with known experimental value
Similarity index = 0.958
Explanation: strongly similar compounds with known experimental value in the training set have been found.

Accuracy of prediction for similar molecules
Accuracy index = 1
Explanation: accuracy of prediction for similar molecules found in the training set is good.

Concordance for similar molecules
Concordance index = 1
Explanation: similar molecules found in the training set have experimental values that agree with the predicted value.

Model's descriptors range check
Descriptors range check = True
Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.

Atom Centered Fragments similarity check
ACF index = 1
Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.

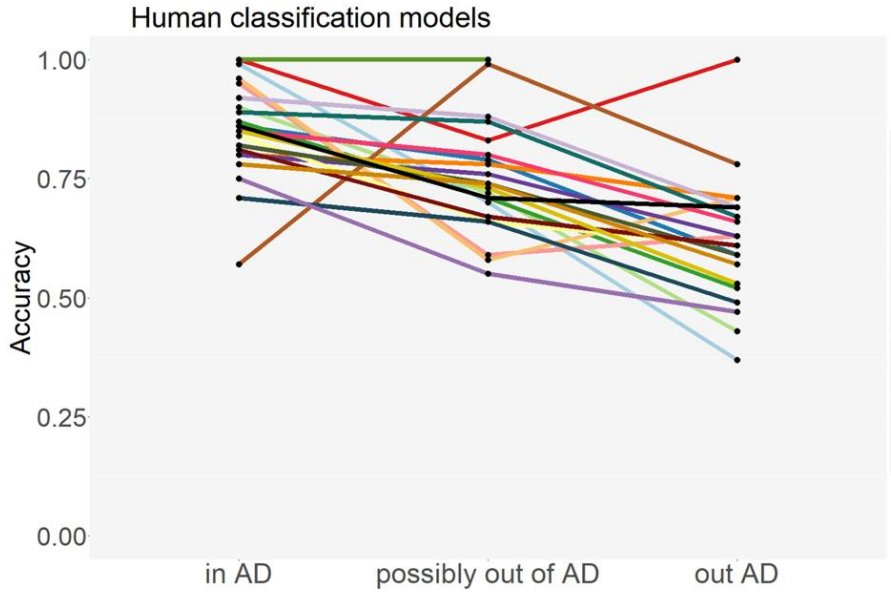
Symbols explanation:

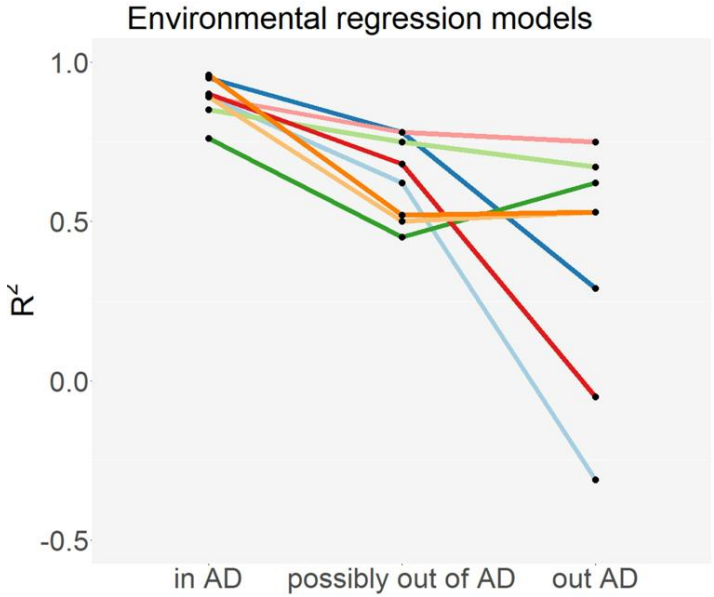
The feature has a good assessment, model is reliable regarding this aspect.

The feature has a non optimal assessment, this aspect should be reviewed by an expert.

The feature has a bad assessment, model is not reliable regarding this aspect.

ADI - The demonstration





Danieli et al. 2023

VERA vs previous tools

VERA's advantages:

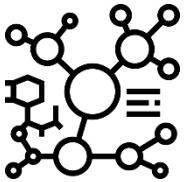
- "Accept almost all" at the beginning
- Multiple metrics
- Memberships
- Comparing clusters
- Predictions
- Batch mode
- Integrated with QSAR

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

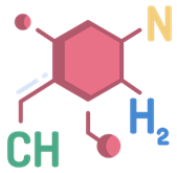
VERA

VERA: the new concept of similarity



STRUCTURAL SIMILARITY

TOXICOLOGICAL SIMILARITY



GROUPING SIMILARITY

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

VERA

General Workflow of VERA

1 ***Calculation of VEGA sim.*** → ***and filtering VEGA sim > 0.65***

2 ***Searching SA in the target*** → ***and filtering by SA***

3 ***Searching MGs in the target*** → ***calculation of grouping similarity*** → ***Considering MG in common and not with the target***

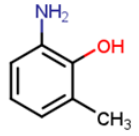
$$GSI = \frac{1+\frac{nMGs \in common}{ntargetMGs}}{2} - \frac{\frac{nMGsNOT \in common}{nsimilarMGs}}{8}$$

4 ***Reasoning*** → ***Clusters SA vs MG_n***

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

VERA

General Workflow of VERA



Target Molecule

VEGA similarity > 0.65

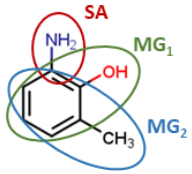
	Similarity	SMILES	Experimental value
0	0.938	COc1ccc(C)cc1N	1
1	0.915	COc1ccccc1N	1
2	0.914	COc1ccc(N)c(C)c1	1
3	0.910	COc1ccccc1[NH3+]	1
4	0.906	COc1ccc(N)cc1	0
...	...	...	...
483	0.651	CC(C)C(=O)Nc1ccc([N+](=O)[O-])c(C(F)(F)F)c1	1
484	0.650	COc1ccc2c3c(cccc13)OC([N+](=O)[O-])=C2	1
485	0.650	Clc1ccccc1	1
486	0.650	CC(C)(C)NCC(O)COc1cccc2c1CCCC2=O	1
487	0.650	COc1ccc(C(=O)C(Br)=CC(=O)O)cc1	1

Searching SA in target

Filter similar compounds with SA

SMILES	VEGA_Sim	Experimental value
Cc1ccccc1N	0.875	1
Cc1cccc(N)c1	0.890	1
Cc1ccc(N)cc1	0.878	1
Nc1ccc(-c2ccccc2)cc1O	0.837	0
COc1ccc(C)cc1N	0.938	1
COc1ccc(N)c(C)c1	0.914	1
Cc1ccc(C)c(N)c1	0.884	1
Cc1cc(-c2ccccc2C)ccc1N	0.742	1
Cc1cc(C)c(N)cc1C	0.848	1
CCn1c2ccccc2c2cc(N)ccc21	0.705	1

Searching MGs in target

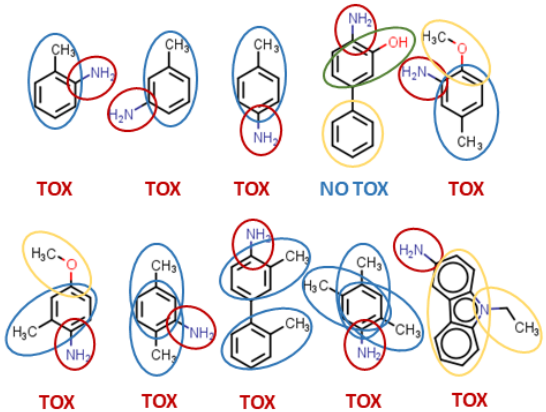


Searching MGs in similar compounds...

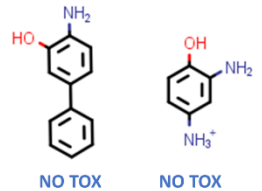
...calculation of grouping similarity

SMILES	VEGA_Sim	Experimental value	GRP_Sim
Cc1ccccc1N	0.875	1	0.857
Cc1cccc(N)c1	0.890	1	0.786
Cc1ccc(N)cc1	0.878	1	0.786
Nc1ccc(-c2ccccc2)cc1O	0.837	0	0.761
COc1ccc(C)cc1N	0.938	1	0.744
COc1ccc(N)c(C)c1	0.914	1	0.744
Cc1ccc(C)c(N)c1	0.884	1	0.744
Cc1cc(-c2ccccc2C)ccc1N	0.742	1	0.744
Cc1cc(C)c(N)cc1C	0.848	1	0.723
CCn1c2ccccc2c2cc(N)ccc21	0.705	1	0.716

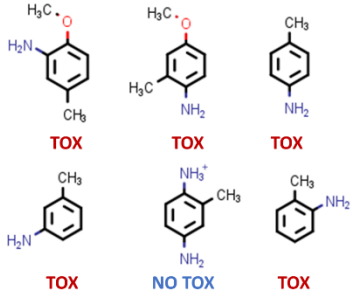
10 most similar compounds



Grouping: SA + MG1



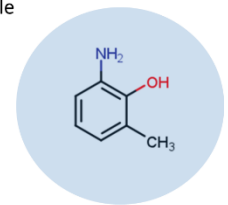
Grouping: SA + MG2



Cluster: INACTIVE

MG1: possible exception rule

VERA ASSESSMENT:



Cluster: ACTIVE

Not Toxic due to exception rules

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

VERA

Pubblication

Open Access Article

Virtual Extensive Read-Across: A New Open-Access Software for Chemical Read-Across and Its Application to the Carcinogenicity Assessment of Botanicals

by Edoardo Luca Viganò [1], Erika Colombo [1], Giuseppa Raitano [1], Alberto Manganaro [2], Alessio Sommovigo [2], Jean Lou CM Dorne [3] and Emilio Benfenati [1,*]

[1] Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, 20156 Milano, Italy

[2] KODE Srl, 56122 Pisa, Italy

[3] Methodological and Scientific Support Unit, European Food Safety Authority, 43126 Parma, Italy

[*] Author to whom correspondence should be addressed.

Molecules **2022**, *27*(19), 6605; **https://doi.org/10.3390/molecules27196605**

Received: 22 August 2022 / Revised: 23 September 2022 / Accepted: 25 September 2022 / Published: 5 October 2022

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

VERA

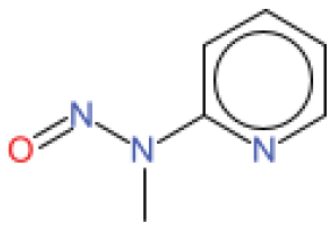
The output

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

VERA - VIRTUAL EXTENSIVE READ ACROSS
Similarity and Grouping for Carcinogenicity endpoint
Input molecule: O=NN(c1ncccc1)C
Vera Group Across - 0.2

VERA - Carcinogenicity model

INPUT MOLECULE



SMILES: O=NN(c1ncccc1)C

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

VERA

The output

1. Prediction Summary

Found in DB: Yes

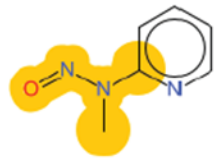
Experimental Value: Active

Prediction	Toxicity	Reliability
Active		

The output

1.1 Structural Alerts in the target molecule

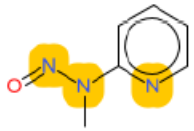
% toxic prevalence in DB

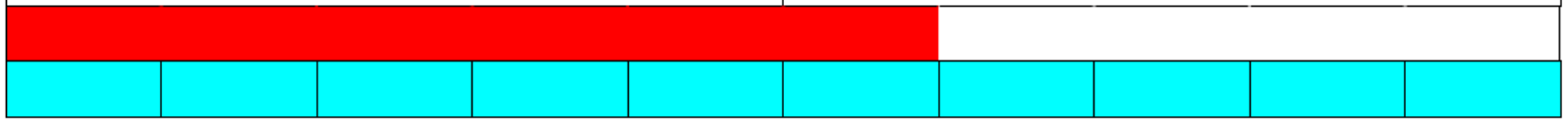
Alert Name	Fragment
SA21 - SA21_gen.Alkyl and aryl N-nitroso groups	

The output

1.2 Molecular Groups in the target molecule

	% toxic prevalence in DB
	Number of compounds used for prediction with MG

Molecular Group name	Fragment
MG1: NH0 - Number of Tertiary amines	



CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

VERA

The output

2. Similar Compounds

2.1 Six most similar compounds

(1)

SMILES: O=NN(c1ncccc1)C

Class: Active

Vega Similarity: 1

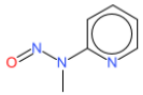
Group Similarity: 1

Similarity Mean: 1

Molecular Groups matches in common with target mol: 5

Molecular Groups type found: 4
* NH0: 2 matches
* aniline: 1 match
* nitroso: 1 match
* pyridine: 1 match

Structural Alerts found: 1
* SA21: 2 matches



(2)

SMILES: O=NN(c1cnccc1)C

Class: Non Active

Vega Similarity: 0.915

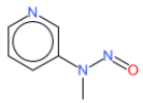
Group Similarity: 1

Similarity Mean: 0.958

Molecular Groups matches in common with target mol: 5

Molecular Groups type found: 4
* NH0: 2 matches
* aniline: 1 match
* nitroso: 1 match
* pyridine: 1 match

Structural Alerts found: 1
* SA21: 2 matches



(3)

SMILES: O=NN(c1ccncc1)C

Class: Non Active

Vega Similarity: 0.916

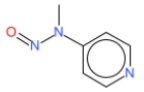
Group Similarity: 1

Similarity Mean: 0.958

Molecular Groups matches in common with target mol: 5

Molecular Groups type found: 4
* NH0: 2 matches
* aniline: 1 match
* nitroso: 1 match
* pyridine: 1 match

Structural Alerts found: 1
* SA21: 2 matches



(4)

SMILES: O=NN(c3ccc(C=Cc2ccnc1ccccc12)cc3)C

Class: Active

Vega Similarity: 0.689

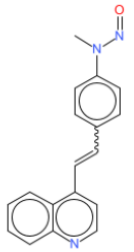
Group Similarity: 0.953

Similarity Mean: 0.821

Molecular Groups matches in common with target mol: 5

Molecular Groups type found: 7
* NH0: 2 matches
* aniline: 1 match
* nitroso: 1 match
* pyridine: 1 match
* bicyclic: 1 match
* benzene: 1 match
* para_hydroxylation: 1 match

Structural Alerts found: 1
* SA21: 2 matches



(5)

SMILES: O=NN2CCCCC2(c1cnccc1)

Class: Active

Vega Similarity: 0.732

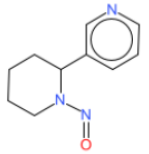
Group Similarity: 0.9

Similarity Mean: 0.816

Molecular Groups matches in common with target mol: 4

Molecular Groups type found: 3
* NH0: 2 matches
* nitroso: 1 match
* pyridine: 1 match

Structural Alerts found: 1
* SA21: 2 matches



(6)

SMILES: O=NN2CC=CCC2(c1cnccc1)

Class: Non Active

Vega Similarity: 0.741

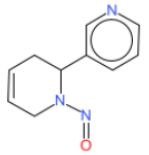
Group Similarity: 0.9

Similarity Mean: 0.821

Molecular Groups matches in common with target mol: 4

Molecular Groups type found: 3
* NH0: 2 matches
* nitroso: 1 match
* pyridine: 1 match

Structural Alerts found: 1
* SA21: 2 matches



CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

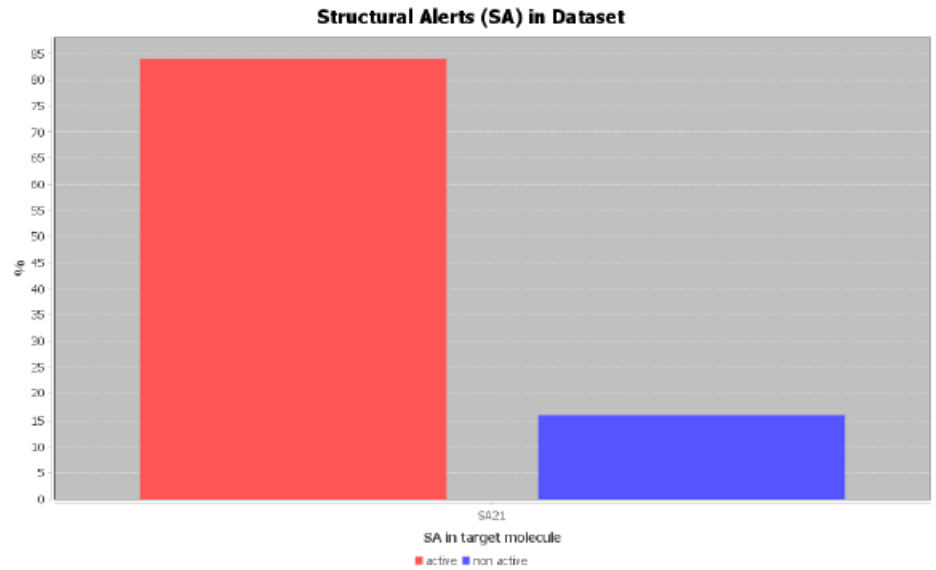
VERA

The output

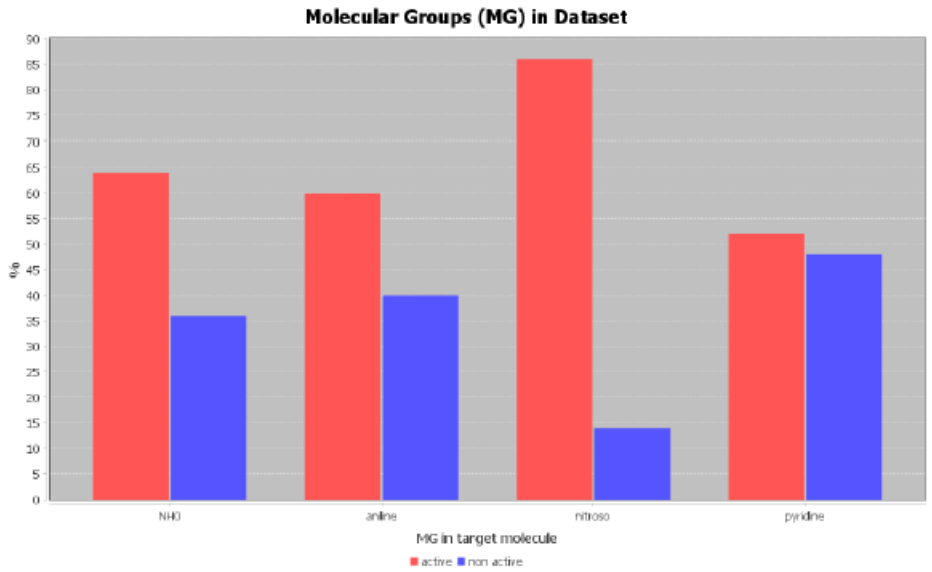
2.2 Toxic Prevalence of Structural Alerts and Molecular Groups

Analysis of the toxic/non-toxic prevalence of the structural alerts and Molecular Groups found in the target in the all dataset.

Target's Structural Alerts prevalence in DB



Target's Molecular Groups prevalence in DB



CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

VERA

The output

3. Reasoning

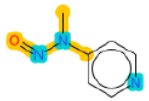
In this section , clusters with the SA with the Molecular Groups are shown

Cluster 1

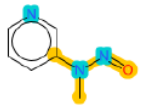
Ortogonal research based on presence of Benigni Bossa structural alerts SA21 and molecular group NH0

This cluster is labeled **Active**

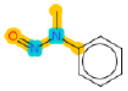
(1)
SMILES: O=NN(c1ccncc1)C
Class: Non Active
Vega Similarity: 0.916



(2)
SMILES: O=NN(c1cnccc1)C
Class: Non Active
Vega Similarity: 0.915

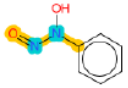


(3)
SMILES: O=NN(c1ccccc1)C
Class: Active
Vega Similarity: 0.844



(4)
SMILES: O=NN([O-])c1ccccc1
Class: Active
Vega Similarity: 0.811

(5)
SMILES: O=NN(O)c1ccccc1
Class: Active
Vega Similarity: 0.811



The output

4. Cluster's results

Cluster 1

SA21 and molecular group NH0

This cluster is labeled Active

Cluster 2

SA21 and molecular group aniline

This cluster is labeled Active

Cluster 3

SA21 and molecular group nitroso

This cluster is labeled Active

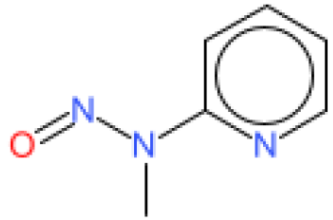
Cluster 4

SA21 and molecular group pyridine

This cluster is labeled Active

According to these clusters, the final assessment of target molecule is :

Active



CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

SWAN

Software implementing Weight of evidence for Assessing the properties of substances, integrating Non-testing method

Integrates

VERA

QSAR Workflow (JANUS)

Reliability of WoE prediction

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

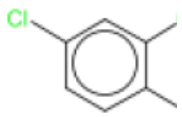
LIFE17 GIE/IT/000461

Some Examp

SWAN

SWAN - Carcinogenicity model

INPUT MOLECULE



SMILES: C1=CC(=CC=C1(

1. Prediction Summary

Found in DB: No
Experimental Value: Not found

Prediction made by the applica

Prediction	Toxicity	Reliability
Carcinogenic		

2. QSAR Workflow - details

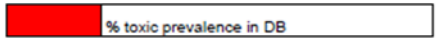
QSAR Workflow prediction: Carcinogenic
QSAR Workflow reliability: Moderate Reliability (0.6)

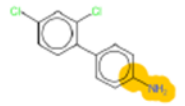
Model name	M
Carcinogenicity classification model (Antares) prediction	Carcinogen (M
Carcinogenicity classification model (Benigni-Bossa rulebase) prediction	Carcinogen (G
Carcinogenicity classification model (Caesar) prediction	NON-Carcinog
Carcinogenicity classification model (ISSCAN-CGX) prediction	Carcinogen (G
Carcinogenicity classification model (Oral - IRFMN) prediction	Carcinogen (G
Carcinogenicity classification model (Inhalation - IRFMN) prediction	Carcinogen (G
Carcinogenicity slope factor model (Oral - IRFMN) prediction	-0.07 (GOOD r
Carcinogenicity slope factor model (Inhalation - IRFMN) prediction	1.1 (MODERA

VERA Read Across tool: Carcinogenic - High Reliability
QSAR Workflow: Carcinogenic - Moderate Reliability

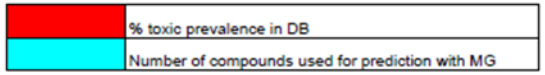
3 VERA Read Accross

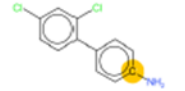
3.1 Structural Alerts in the target molecule

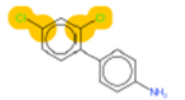


Alert Name	Fragment
SA28 - SA28_gen.Primary aromatic amine, hydroxyl amine and its derived esters (with restrictions). SA28 Primary aromatic amine, hydroxyl amine and its derived esters (with restrictions): Primary aromatic amine, hydroxyl amine and its derived esters (with	

3.2 Molecular Groups in the target molecule

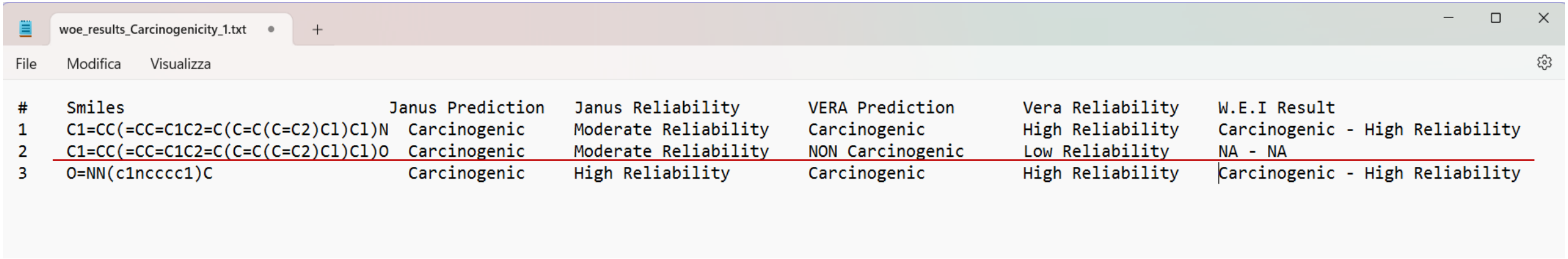


Molecular Group name	Fragment
MG1: ArN - Number of N functional groups attached to aromatics	

Molecular Group name	Fragment
MG3: Ar - Cl, Br bound to aromatic ring	

Solving conflicts

Based on the relative **reliability** of the QSAR versus Read-Across


woe_results_Carcinogenicity_1.txt

File Modifica Visualizza

#	Smiles	Janus Prediction	Janus Reliability	VERA Prediction	Vera Reliability	W.E.I Result
1	C1=CC(=CC=C1C2=C(C=C(C=C2)Cl)Cl)N	Carcinogenic	Moderate Reliability	Carcinogenic	High Reliability	Carcinogenic - High Reliability
2	C1=CC(=CC=C1C2=C(C=C(C=C2)Cl)Cl)O	Carcinogenic	Moderate Reliability	NON Carcinogenic	Low Reliability	NA - NA
3	O=NN(c1ncccc1)C	Carcinogenic	High Reliability	Carcinogenic	High Reliability	Carcinogenic - High Reliability

Conclusions

- A conceptual scheme to replicate the expert's approach
- A tool to integrate QSAR and read-across
- Improved reasoning on both QSAR and read-across
- General methods and endpoint-specific components

Next steps

Extending to more endpoints
Further parameters to be added

Thank you for your attention!

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS