THE USE OF VEGA AND JANUS SOFTWARE TO EVALUATE CARCINOGENICITY OF BOTANICALS

EMILIO BENFENATI, ERIKA COLOMBO, EDOARDO VIGANÒ

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI

MILANO, ITALY

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

EFSA Guidance on WoE

Approach for WoE

1. Gather all info
2. Evaluate individual lines of evidence
3. Integrate the results

EFSA Guidance: integration

Criteria for integration

1. Relevance
2. Reliability
3. Agreement

In silico and read-across: integration

Environment International 131 (2019) 105060

Contents lists available at ScienceDirect

Environment International

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

Review article

Integrating *in silico* models and read-across methods for predicting toxicity of chemicals: A step-wise strategy

Emilio Benfenati[a,*], Qasim Chaudhry[b], Giuseppina Gini[c], Jean Lou Dorne[d]

[a] *Department of Environmental and Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via La Masa 19, Milano, Italy*
[b] *University of Chester, Parkgate Road, Chester CH1 4BJ, United Kingdom*
[c] *Politecnico di Milano, piazza L. da Vinci 32, Milano, Italy*
[d] *Scientific Committee and Emerging Risks Unit, European Food Safety Authority, Via Carlo Magno 1A, Parma, Italy*

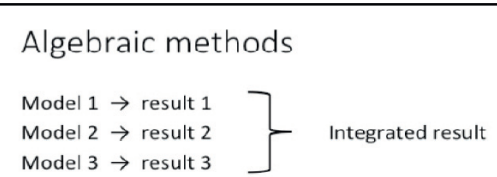
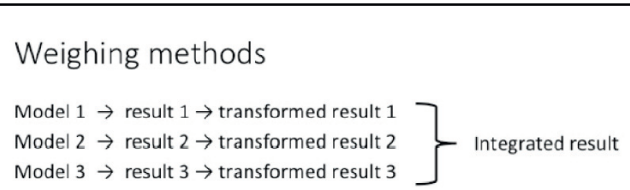
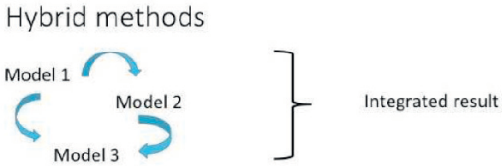
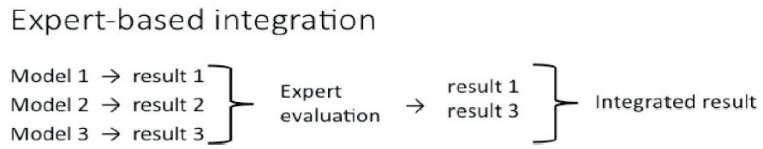
ARTICLE INFO

Handling Editor: Da Chen

ABSTRACT

In silico methods and models are increasingly used for predicting properties of chemicals for hazard identification and hazard characterisation in the absence of experimental toxicity data. Many *in silico* models are available and can be used individually or in an integrated fashion. Whilst such models offer major benefits to toxicologists, risk assessors and the global scientific community, the lack of a consistent framework for the integration of *in silico* results can lead to uncertainty and even contradictions across models and users, even for the same chemicals. In

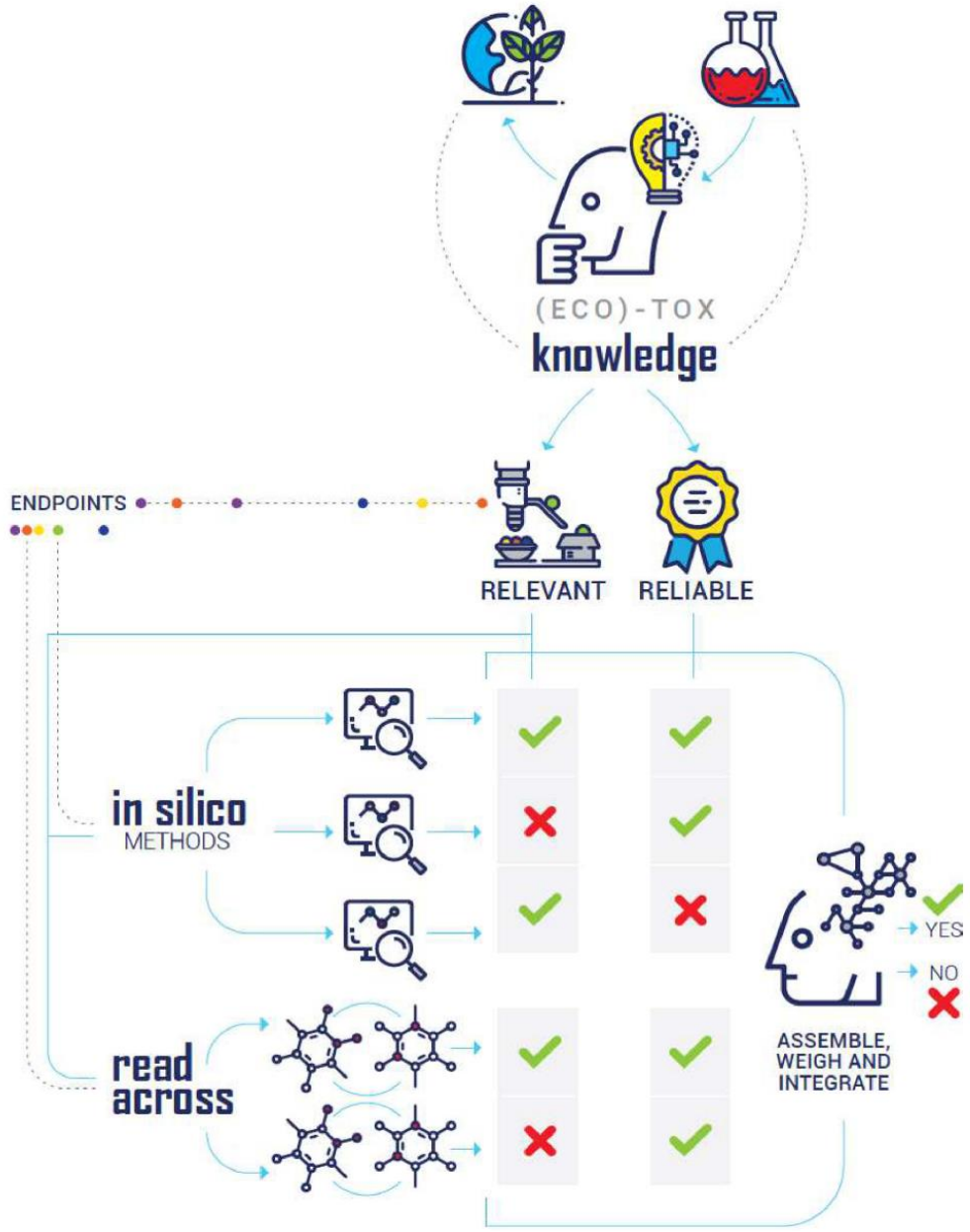
Integration of in silico

Algebraic and voting methods	
Weighing	
Hybrid	
Learning	
Expert-based	

Example of weighing

$$\mathrm{CONSENSUS} = \frac{(\pm 1) * \mathrm{AD}_{\mathrm{CAESAR}} + (\pm 1) * \mathrm{AD}_{\mathrm{SARPY}} + (\pm 1) * \mathrm{AD}_{\mathrm{TTVEGA}}}{\mathrm{AD}_{\mathrm{CAESAR}} + \mathrm{AD}_{\mathrm{SARPY}} + \mathrm{AD}_{\mathrm{TTVEGA}}}$$

Integrating in silico and read-across



Use all lines of evidence

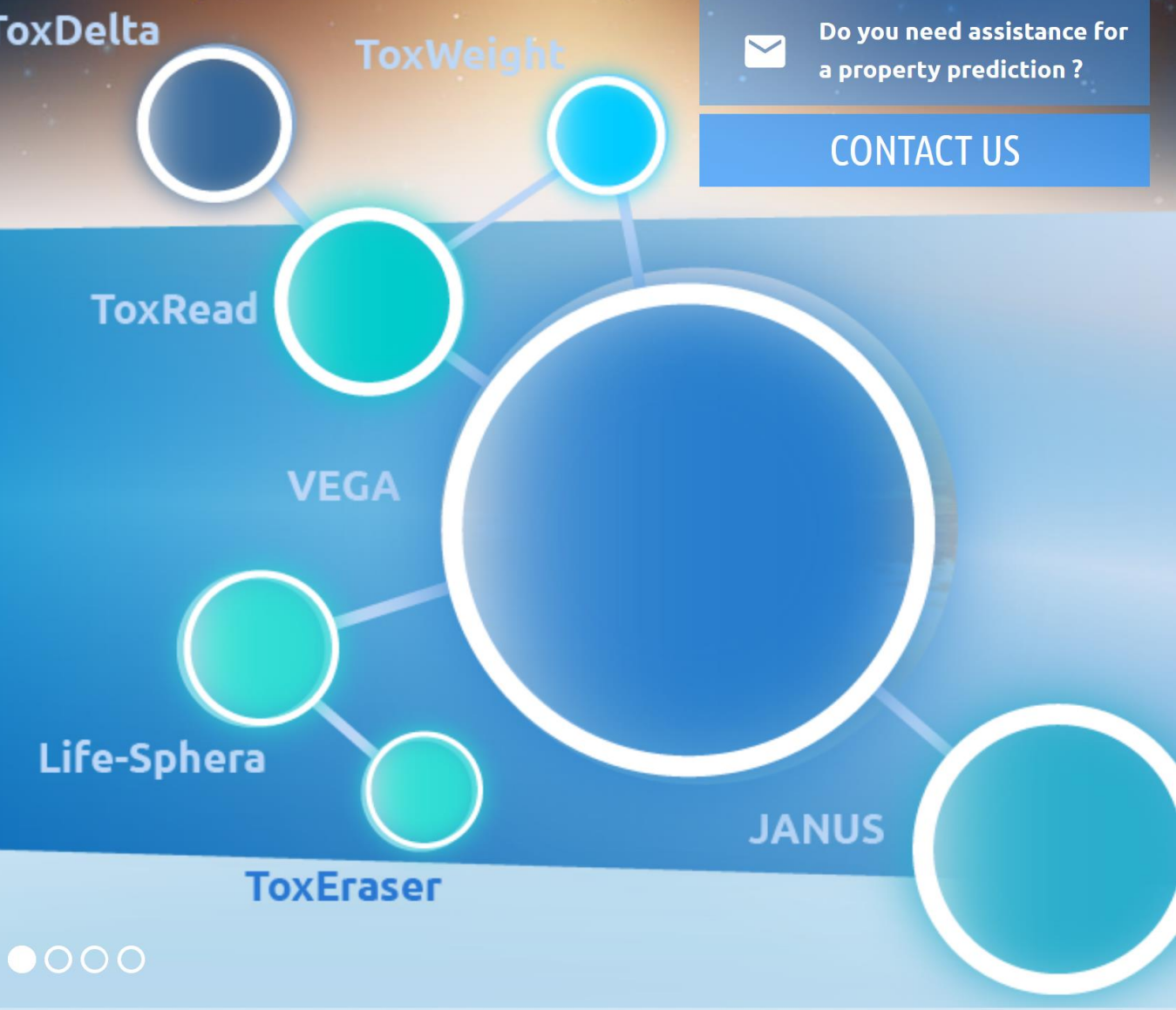
1. VEGA in silico models
2. Read-across
3. Reasoning

- Check agreement

Welcome to the VEGA HUB

Offering a family of tools to evaluate chemical hazard: VEGA, ToxRead, ToxWeight, ToxDelta, and JANUS.

VEGA is the QSAR software with tens of models for individual properties.



Do you need assistance for a property prediction ?

CONTACT US

https://www.vegahub.eu/

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.

All VEGA models provide an applicability domain index (ADI) value based on several indices

3.2 Applicability Domain:
Measured Applicability Domain Scores

Global
AD ind
Explan

Simila
Similar
Explan

Accura
Accura
Explan

Concc
Concc
Explan
value.

Model
Descri
Explan
training

Atom
ACF in
Explan
set.

Symbols expla

The fea

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

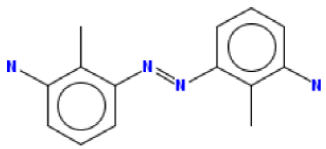
The feature has a bad assessment, model is not reliable regarding this aspect.

VEGA

Mutagenicity (Ames test) model (CAESAR) 2.1.13 page 1

1. Prediction Summary

Prediction for compound Molecule 0

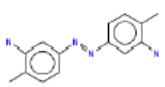


Prediction: Reliability:

Prediction is Mutagenic, the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections.
The following relevant fragments have been found: SA29 Aromatic diazo

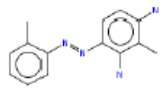
Compound: Molecule 0
Compound SMILES: N(=Nc1cccc(N)c1C)c2cccc(N)c2C
Experimental value: -
Predicted Mutagen activity: Mutagenic
Structural alerts: SA29 Aromatic diazo
Reliability: the predicted compound is into the Applicability Domain of the model
Remarks:
none

3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



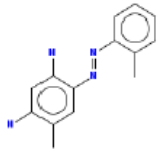
CAS: 119631-00-4
Dataset id: 3482 (Training set)
SMILES: N(=Nc1ccc(c(N)c1)C)c2ccc(c(N)c2)C
Similarity: 0.977

Experimental value: Mutagenic
Predicted value: Mutagenic



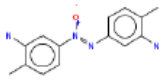
CAS: 84434-45-7
Dataset id: 2696 (Test set)
SMILES: N(=Nc1ccccc1C)c2ccc(N)c(c2(N))C
Similarity: 0.951

Experimental value: Mutagenic
Predicted value: Mutagenic



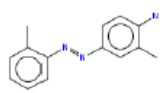
CAS: 7467-29-0
Dataset id: 3021 (Training set)
SMILES: N(=Nc1cc(c(N)cc1(N))C)c2ccccc2C
Similarity: 0.949

Experimental value: Mutagenic
Predicted value: Mutagenic



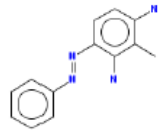
CAS: 5857-92-1
Dataset id: 2566 (Training set)
SMILES: [O-][N+](=Nc1ccc(c(N)c1)C)c2ccc(c(N)c2)C
Similarity: 0.93

Experimental value: Mutagenic
Predicted value: Mutagenic



CAS: 97-56-3
Dataset id: 594 (Training set)
SMILES: N(=Nc1ccccc1C)c2ccc(N)c(c2)C
Similarity: 0.925

Experimental value: Mutagenic
Predicted value: Mutagenic



CAS: 84434-42-4
Dataset id: 1799 (Training set)
SMILES: N(=Nc1ccc(N)c(c1(N))C)c2ccccc2
Similarity: 0.924

Experimental value: Mutagenic
Predicted value: Mutagenic

ADI parameters

- ✓ **Visualization of similar substances**
- ✓ **Similarity index** *(chemical; sub-indices)*
- ✓ **Chemometric check** *(descriptor space)*
- ✓ **Atom centered-fragment** *(chemical)*

Chemical input space

- ✓ **Check of the descriptor sensitivity** *(algorithm)*
- ✓ **Uncertainty** *(algorithm)*

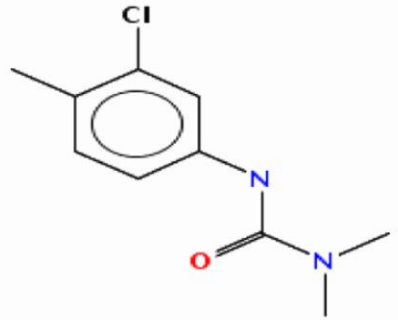
Characteristics of the algorithm

- ✓ **Fragments for outliers** *(output space)*
- ✓ **Prediction Accuracy** *(output space)*
- ✓ **Prediction Concordance** *(tox) exploration)*

Input and output toxicological space

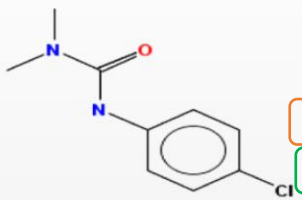
ADI accuracy

Prediction for the compound no. 1: CN(C)C(=O)NC1=CC(Cl)=C(C)C=C1



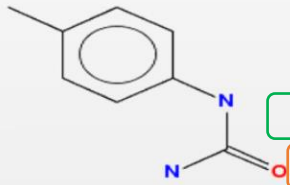
Carcinogenic: Non-Positive
Class indices: Positive=0.079, Non-Positive=0.921
Remarks for the prediction:

The following chemicals similar to the query compound have been identified in the CAESAR database:



Dataset id: 162
SMILES: Clc1ccc(cc1)NC(=O)N(C)C
Similarity: 0.884

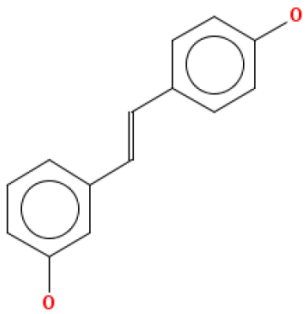
Experimental class: Positive
Predicted class: Non-Positive



Dataset id: 754
SMILES: O=C(Nc1ccc(cc1)C)N
Similarity: 0.857

Experimental class: Non-Positive
Predicted class: Positive

ADI concordance



Prediction: Reliability:

Prediction is Possible NON-Mutagenic, but the result shows some critical aspects, which require to be checked:
- similar molecules found in the training set have experimental values that disagree with the predicted value

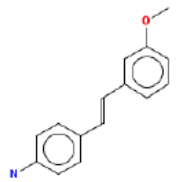
Global AD Index
AD index = 0.719
Explanation: the predicted compound could be out of the Applicability Domain of the model.

Similar molecules with known experimental value
Similarity index = 0.901
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 = 0.33
Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.

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.

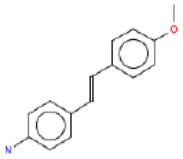


Compound #1

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

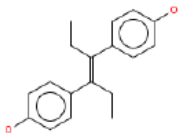


Compound #2

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

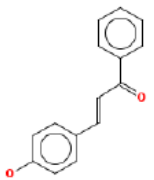


Compound #3

CAS: 56-53-1
Dataset id: 2731 (Test set)
SMILES: Oc1ccc(cc1)C(=C(c2ccc(O)cc2)CC)CC
Similarity: 0.893

Experimental value: NON-Mutagenic
Predicted value: NON-Mutagenic

Alerts (not found in the target): SM158



Compound #4

CAS: 20426-12-4
Dataset id: 561 (Test set)
SMILES: O=C(C=Cc1ccc(O)cc1)c2ccccc2
Similarity: 0.888

Experimental value: NON-Mutagenic
Predicted value: NON-Mutagenic

Alerts (not found in the target): SM158; SM172

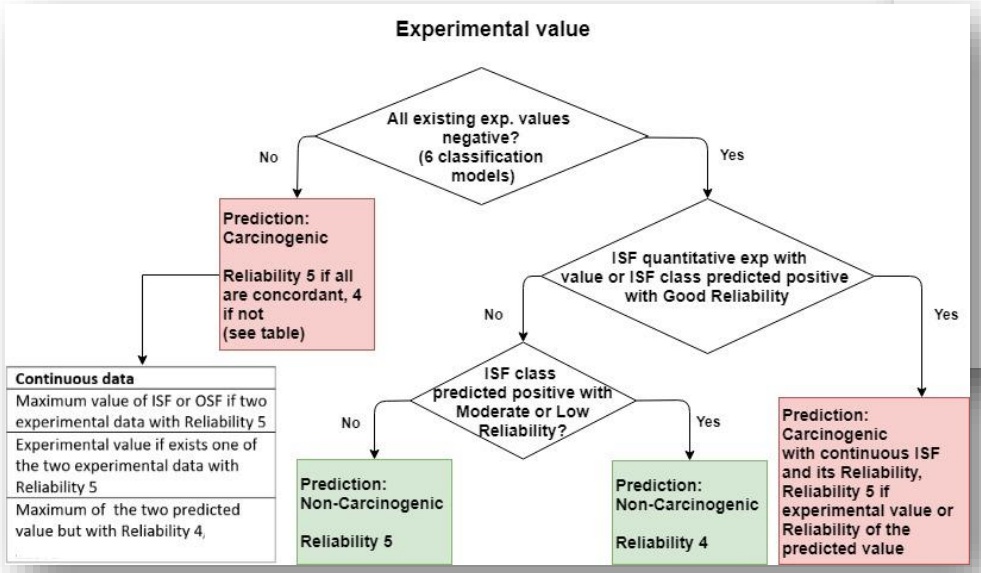
JANUS

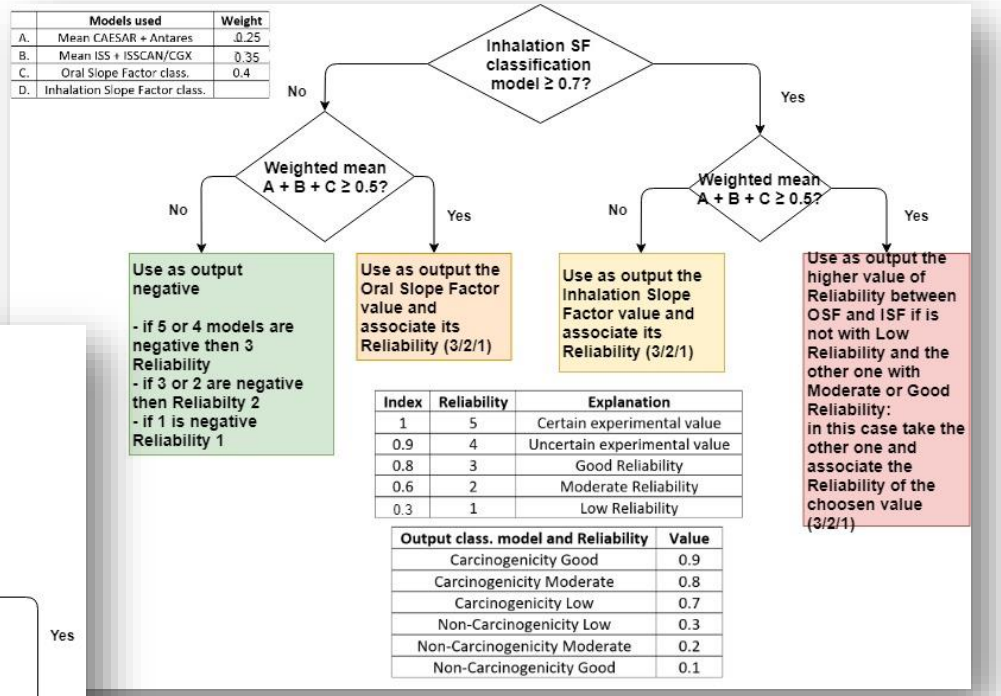
Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit

Umwelt Bundesamt

Carcinogenicity

Oral and inhalation, qualitative and quantitative (slope factor) combined in a single assessment.





	Models used	Weight
A.	Mean CAESAR + Antares	0.25
B.	Mean ISS + ISSCAN/CGX	0.35
C.	Oral Slope Factor class.	0.4
D.	Inhalation Slope Factor class.	

Index	Reliability	Explanation
1	5	Certain experimental value
0.9	4	Uncertain experimental value
0.8	3	Good Reliability
0.6	2	Moderate Reliability
0.3	1	Low Reliability

Output class. model and Reliability	Value
Carcinogenicity Good	0.9
Carcinogenicity Moderate	0.8
Carcinogenicity Low	0.7
Non-Carcinogenicity Low	0.3
Non-Carcinogenicity Moderate	0.2
Non-Carcinogenicity Good	0.1

The case study on botanicals

JANUS: prediction on 80% (mainly non carcino 3:1)

Expert read-across: about 80% (mainly non carcino 2:1)

Agreement between these assessments: 50%

The automatic read-across

Learning from features observed by expert

Role of molecular groups

Multiple similarity:

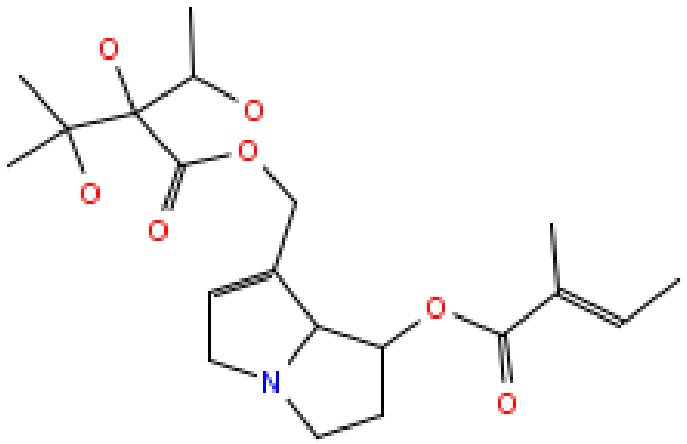
- structural alerts,
- groups,
- overall structural similarity

Solving conflicts

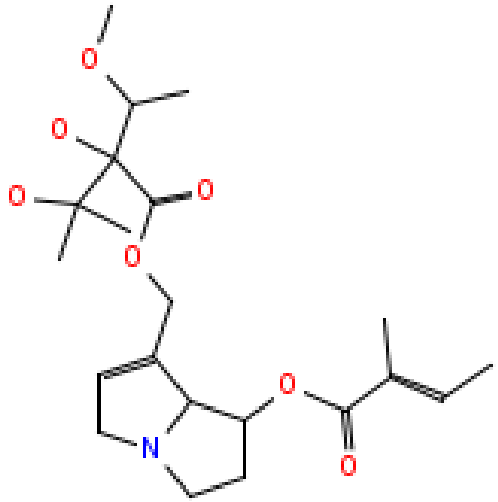
Implementation of software

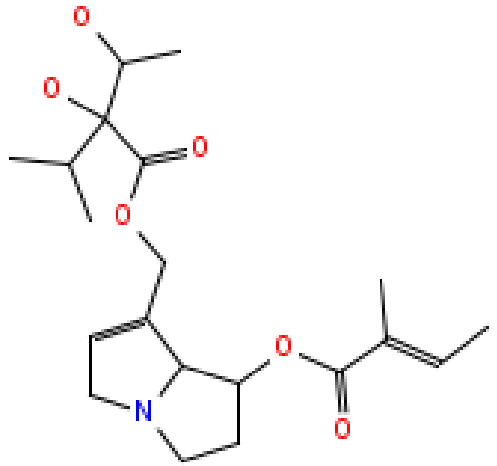
Read-across. Example: toxic

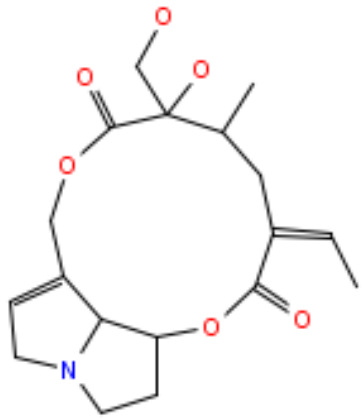
Target: heliosupine

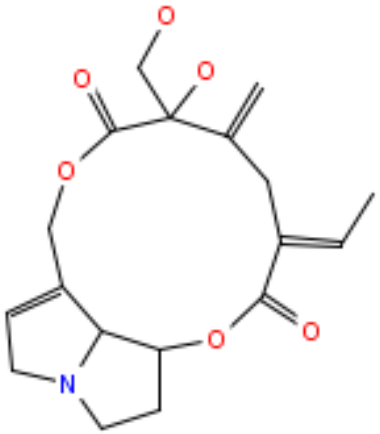


Similar Carcinogen



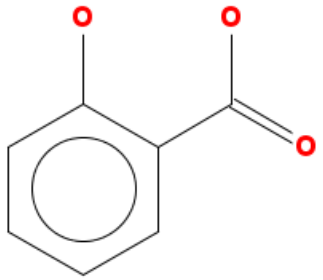




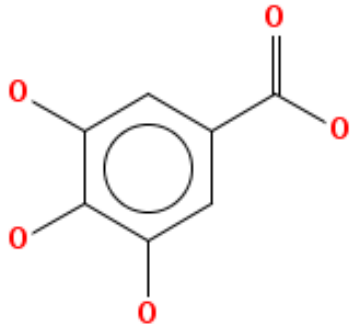


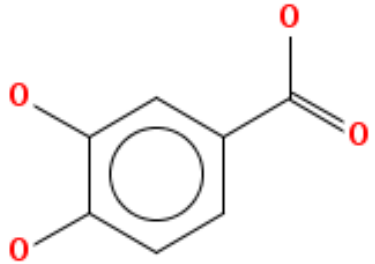
Read-across. Example: non-toxic

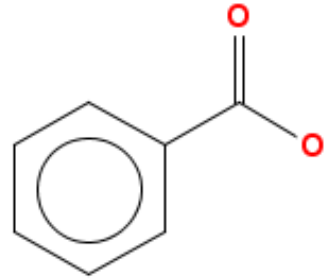
Target (salicylic acid)



Similars





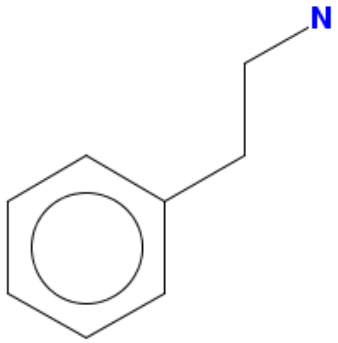


Non carcino

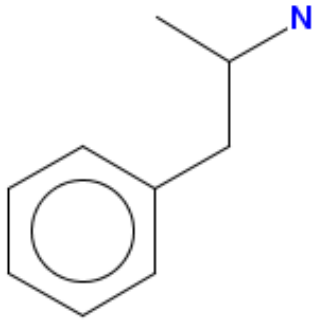
Read-across. Example: ambiguous

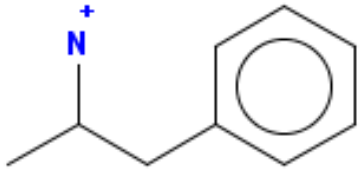
Target.

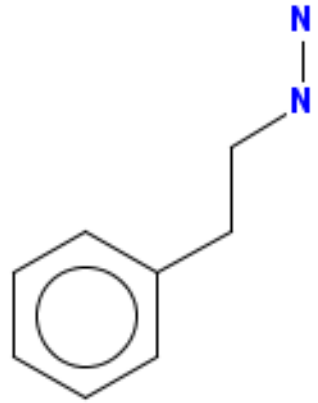
Phenylethylamine



Similars







Non carcino

Carcino

Hydrazine

The new software based on grouping

Target

TEST: Similarity and Grouping for Carcino

Molecule Input:

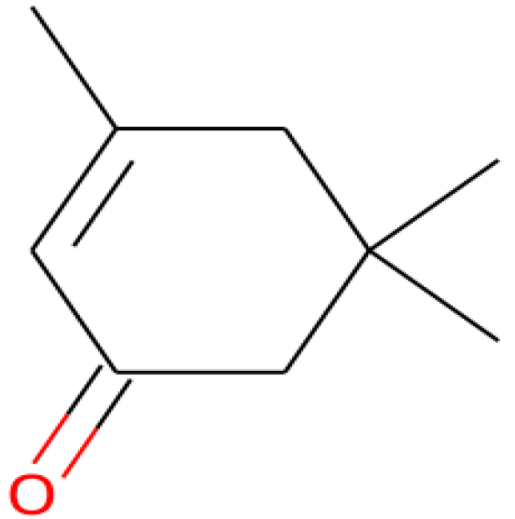
TotGroups: 5

['allylic_oxid', 'ketone', 'CC4', 'ketone_aliphatic']

Alerts:

f_SA10_gen unf_SAturated carbonyls

Stat: Neutro with Prevalence%: 50.0



Experimental Value is known: Carcinogen

Carcino

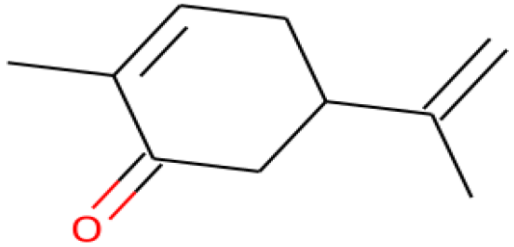
The new software based on grouping (ii)

Similar

1) SMILES: O=C1C(=CCC(C(=C)C)C1)C
Class: NON-Carcinogen
SimilarityVega: 0.92
SimilarityGrouping: 0.838
Grp_Similarity_Mean: 0.879
Groups in common: 4
TotGroups 8

GRUPPI:
allylic_oxid Stat: Neutro
ketone Stat: Neutro
ketone_Topliss Stat: Neutro
CH2_Terminal Stat: Neutro
ketone_aliphatic Stat: Neutro

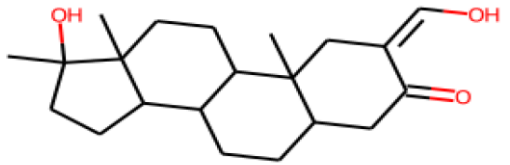
Simil:



Alerts: ['f_SA10']

2) SMILES: O=C1C(=CO)CC4(C)(C(C1)CCC3C4(CCC2(C)(C3(CCC2(O)(C))))
Class: Carcinogen
SimilarityVega: 0.657
SimilarityGrouping: 0.8
Grp_Similarity_Mean: 0.728
Groups in common: 4
TotGroups 15

GRUPPI:
Al_OH Stat: Neutro
allylic_oxid Stat: Neutro
bicyclic Stat: Neutro
ketone Stat: Neutro
ketone_Topliss Stat: Neutro
steroid Stat: Tox
CC4 Stat: Neutro
ketone_aliphatic Stat: Neutro



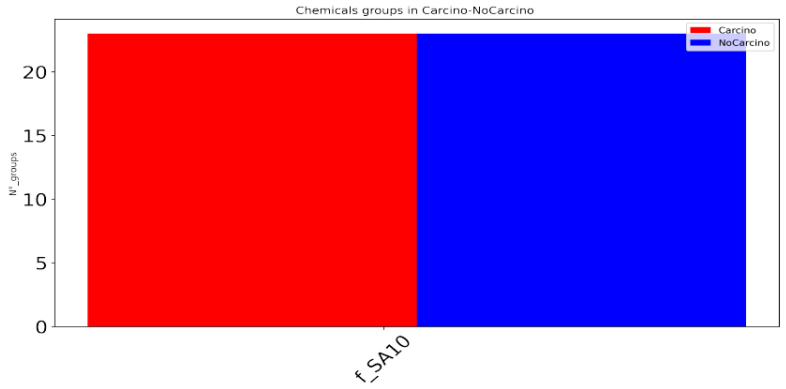
Alerts: ['f_SA10']

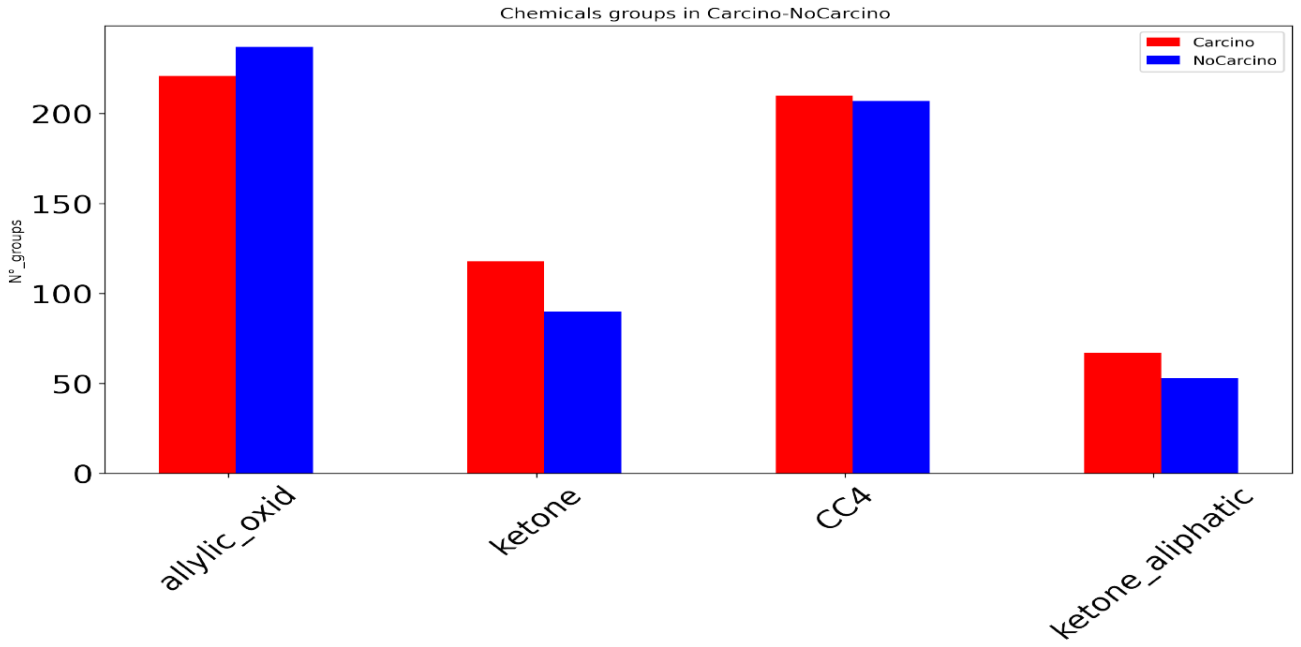
The new software based on grouping (iii)

Structural alert +

molecular groups, screened individually

Molecule Input: Bar Chart for groups in dataset find in target





The new software based on grouping (iv)

Target

TEST: Similarity and Grouping for Carcino

Molecule Input:

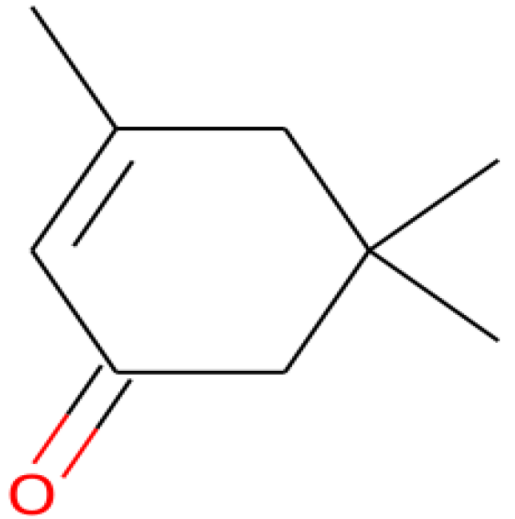
TotGroups: 5

['allylic_oxid', 'ketone', 'CC4', 'ketone_aliphatic']

Alerts:

f_SA10_gen unf_SAturated carbonyls

Stat: Neutro with Prevalence%: 50.0



Experimental Value is known: Carcinogen

Read-across: Carcino

Experimental: Carcino

Statistics on Botanicals

63 botanicals have been evaluated with grouping

10 substances are not assigned because recognized by the tool as ***equivocal***

Confusion Matrix

Real \ Prediction	Carcino	NON-Carcino	Equivocal
Carcino	18	8	4
NON-Carcino	8	19	6

Sensitivity	69.23%
Specificity	70.37%
Accuracy	69.23%
Precision	69.81%

Conclusions

- Carcinogenicity: complex endpoint, with relatively low number of data, experimental uncertainty
- Necessary to screen multiple lines of evidence
- WoE: read-across, reasoning, in silico results
- VEGA provides multiple models (also for read-across & reasoning)
- JANUS integrates results
- New software for read-across based on relevant molecular groups, to be used for screening of these moieties