How to use the VEGAHUB tools to assess fish acute toxicity

Anna Lombardo

Istituto di Ricerche Farmacologiche Mario Negri IRCCS

Disclaimer: This presentation reflects only the author's view, and the Istituto di Ricerche Farmacologiche Mario Negri is not responsible for any use that may be made of the information it contains herein.

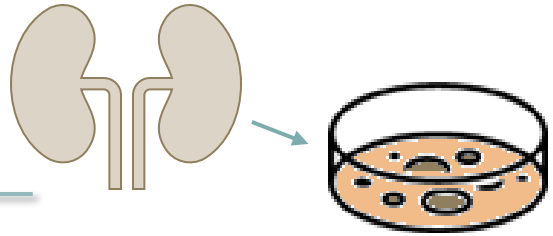
ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

NAMs

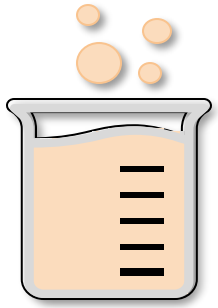
(New Approach Methodologies)

To reduce the dependency on animal testing for the chemical hazard and risk assessment

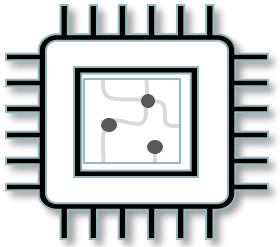
In vitro



Ex vivo

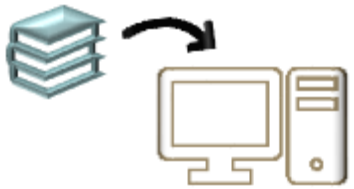


In chemico



In silico

The *in silico* methods



Take advantage of existing data

Can be used to fill the data-gaps

Cheap

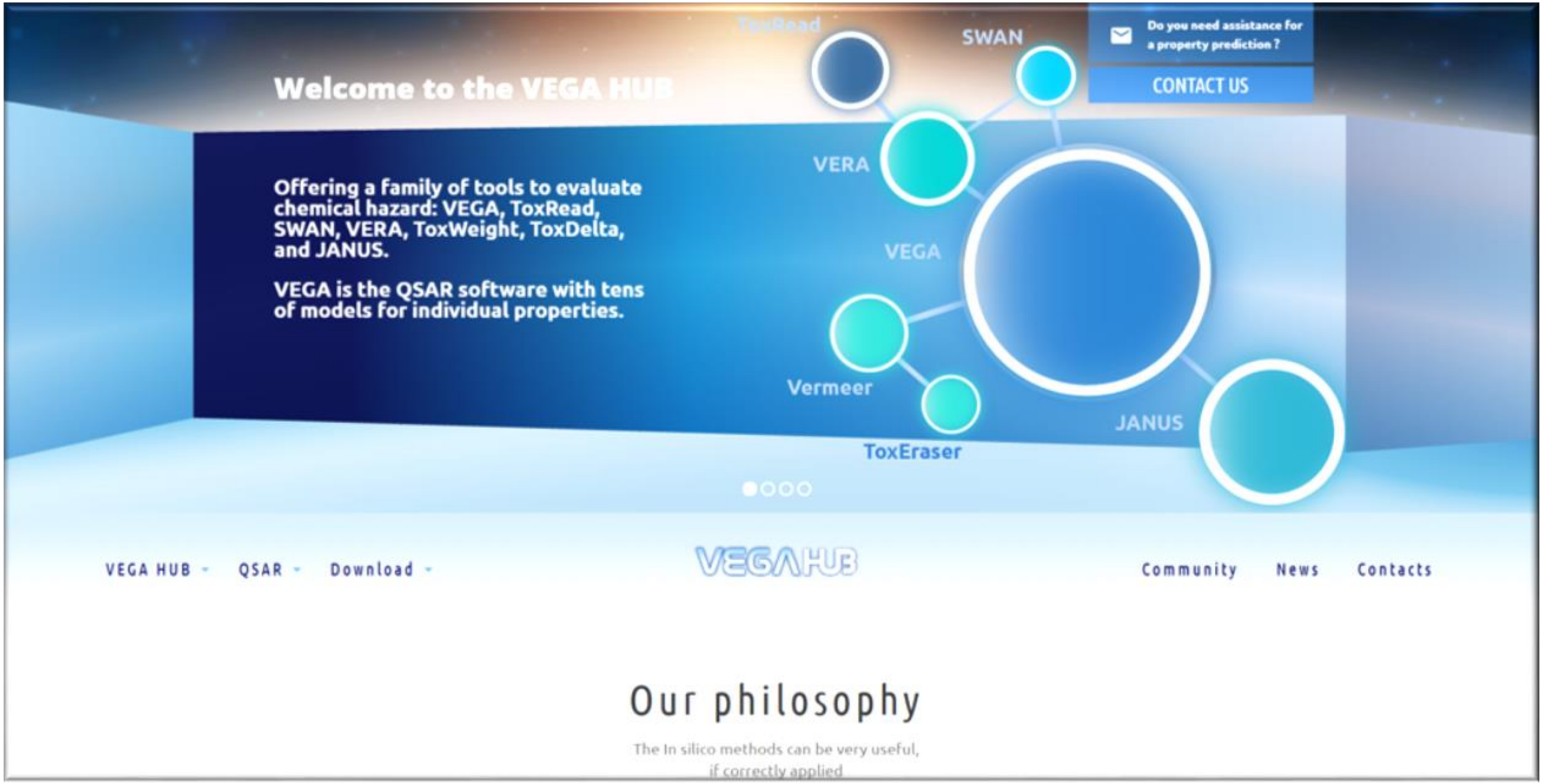
Fast

Batch processing

Green (no solvent, no waste)

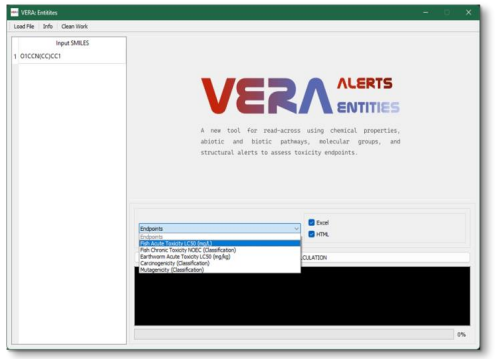
Ethical issues: (no lab. animals)

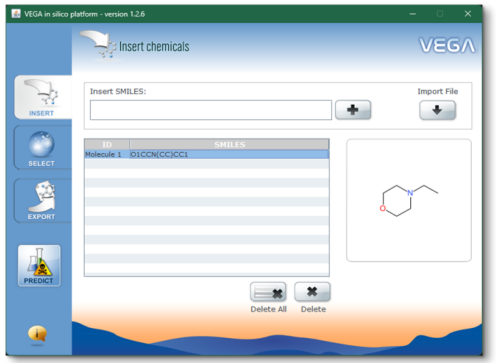
Pre-synthesis studies

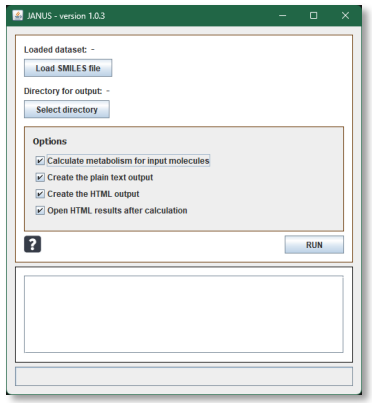


https://www.vegahub.eu/

The VEGAHUB tools: general information







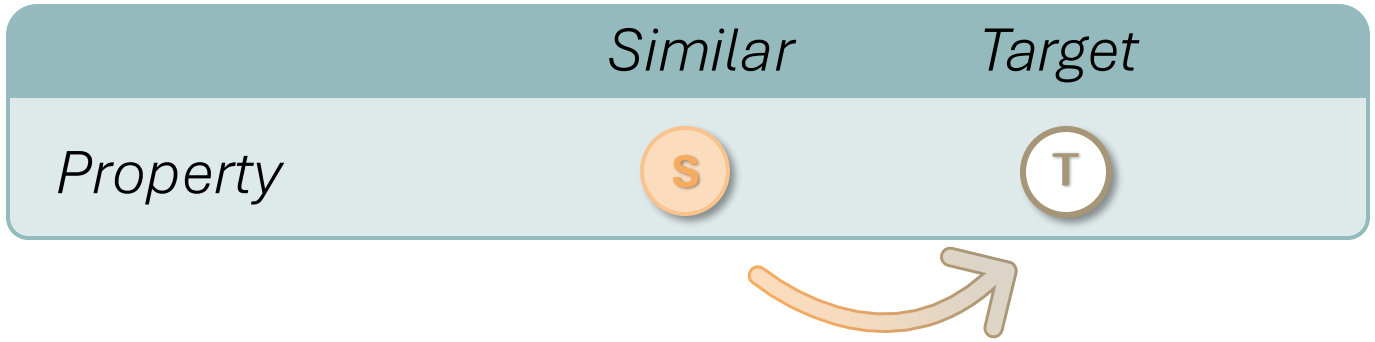
- **User-friendly** tools
- **Transparent** and **information-rich outputs**
- **Goal**: **help**, not substitute assessors

- **Input**: chemical structure (SMILES*)
- **Output**: **tabular** (for batch) + **PDF/HTML** (detailed)

**Alphanumeric strings to represent the molecule (atoms, bonds,…)*

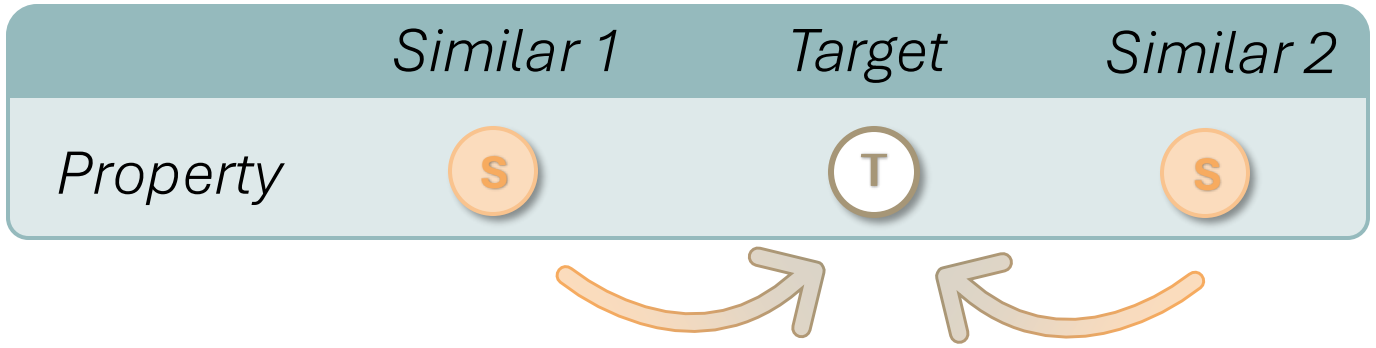
Read-across - Analogue approach

⇨ estimates a property/activity from **similar chemical(s)**

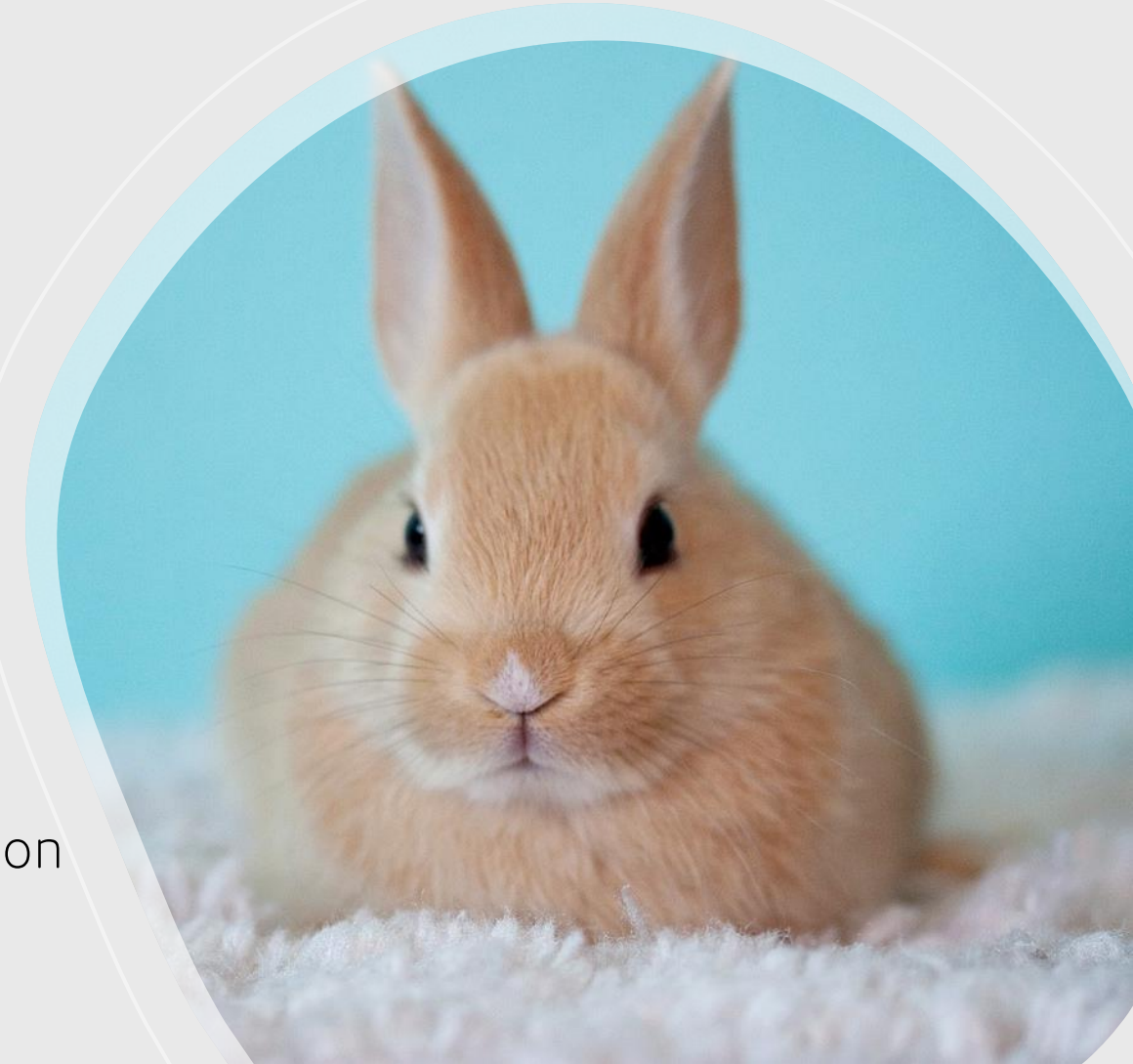


One-to-one

Interpolation (extrapolation)

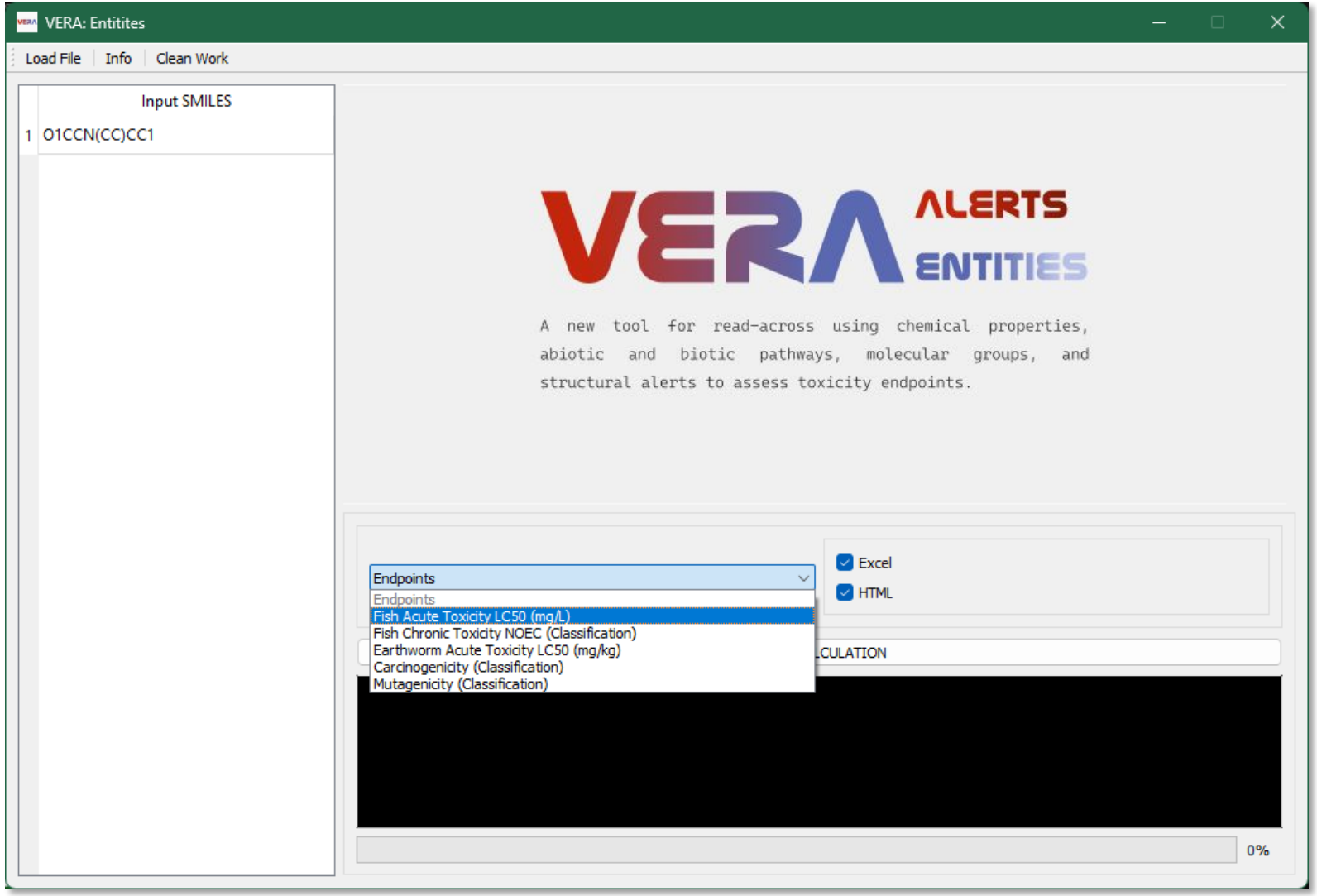


Read-across & similarity



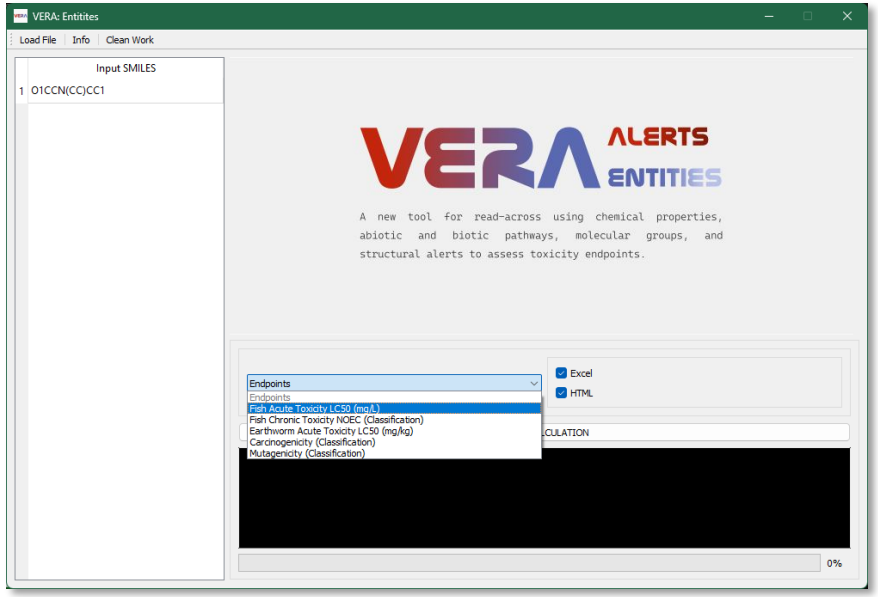
- No accepted definition
- Expert-dependent

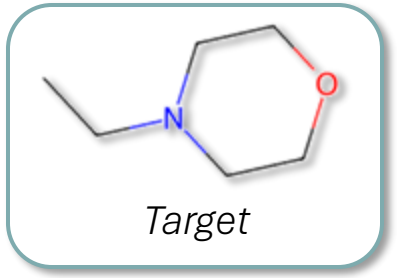
Virtual Extensive Read-Across: VERA

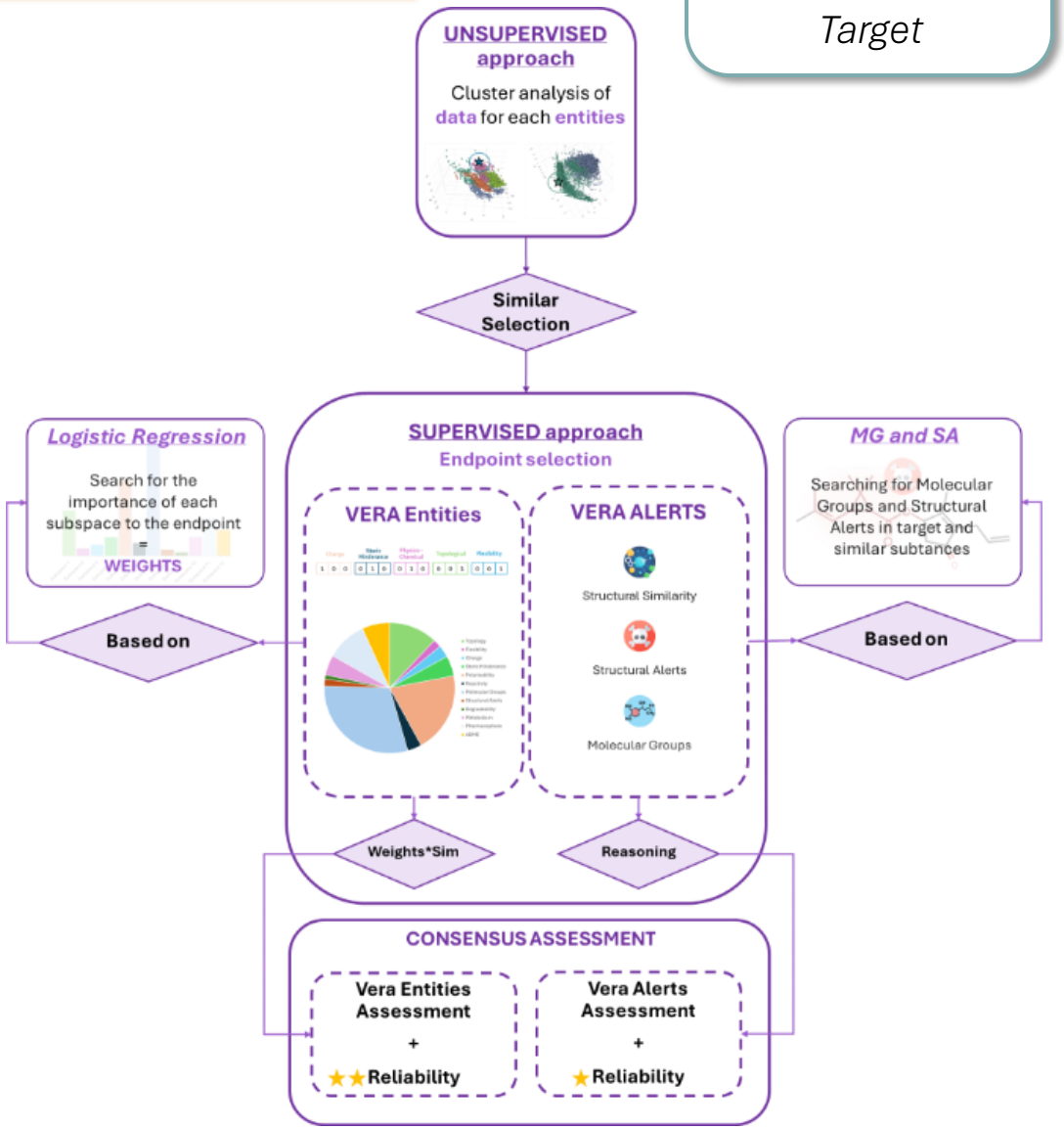


- Automated
- Transparent
- Reproducible
- Unique system with multiple similarity approaches
- Mathematical definition of reliability

Virtual Extensive Read-Across: VERA





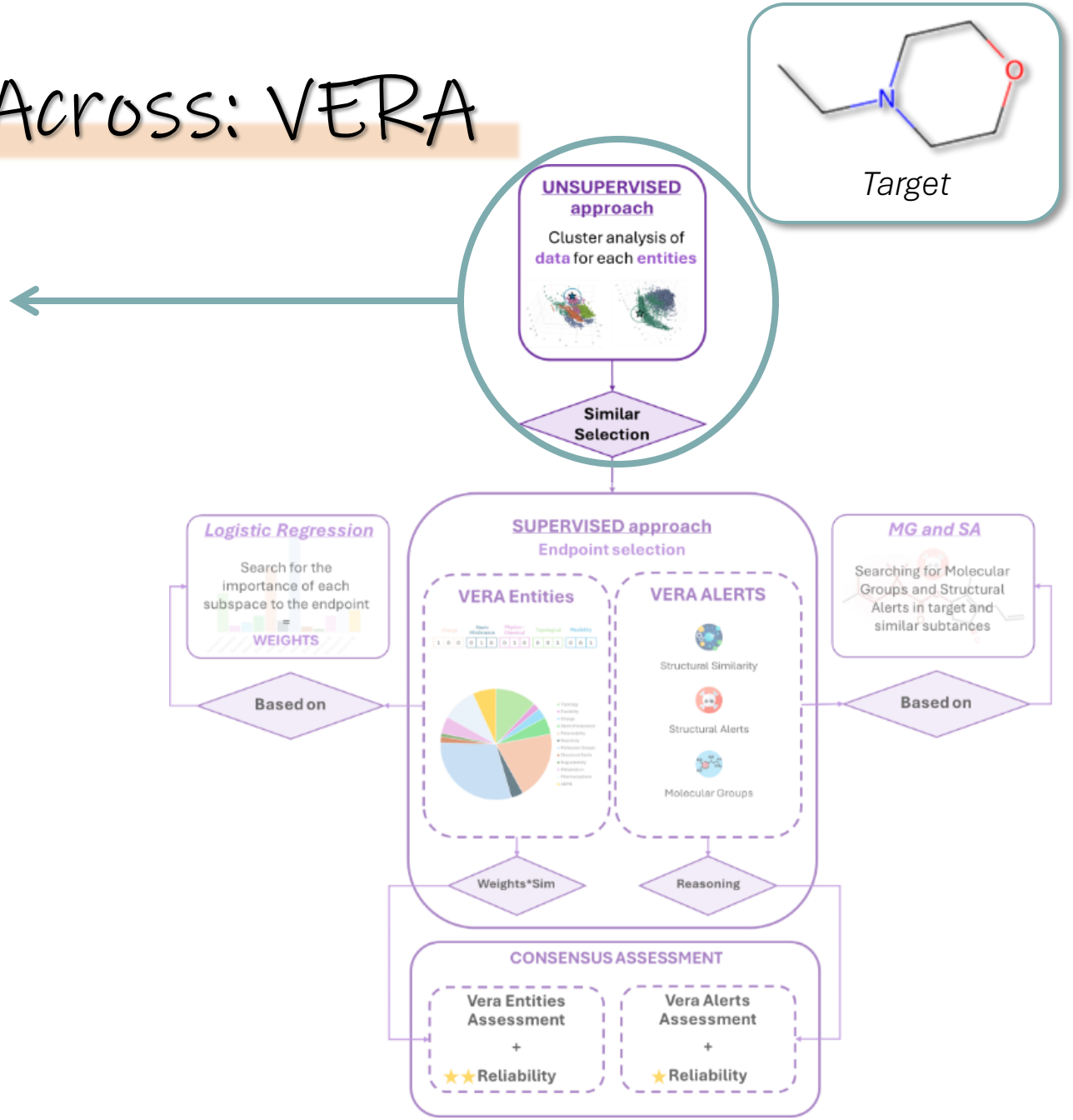


Virtual Extensive Read-Across: VERA

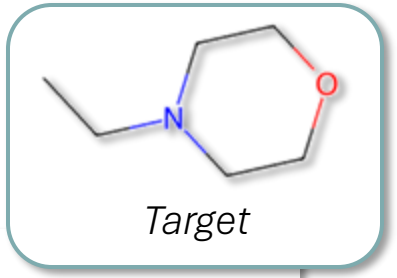
Goal: **similar selection**

Based on:

- 12 **entities** (chemical, biotic, abiotic, physico-chemical)
- 10k chemicals
- **Clusters**



Virtual Extensive Read-Across: VERA



Target

UNSUPERVISED

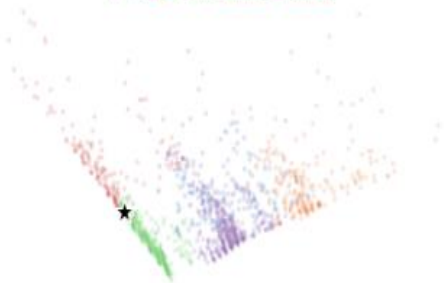
MolecularGroupsSpace

TopologicalSpace

ReactiveSpace

AdmeSpace

PhysChemSpace



PolarizabilitySpace

StericHinderenceSpace

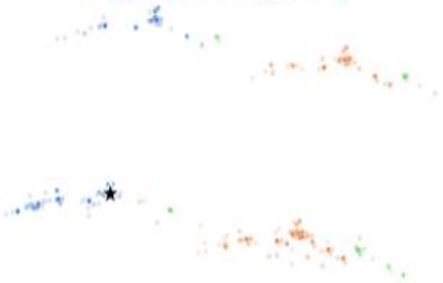
PharmacophoreSpace

ChargeDistributionSpace

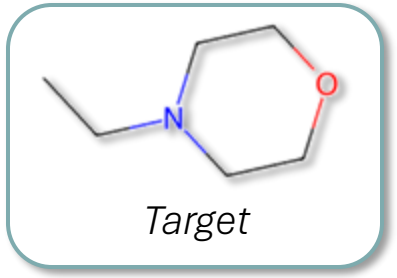
MetaboliteSpace

FlexibilitySpace

DegradabilitySpace

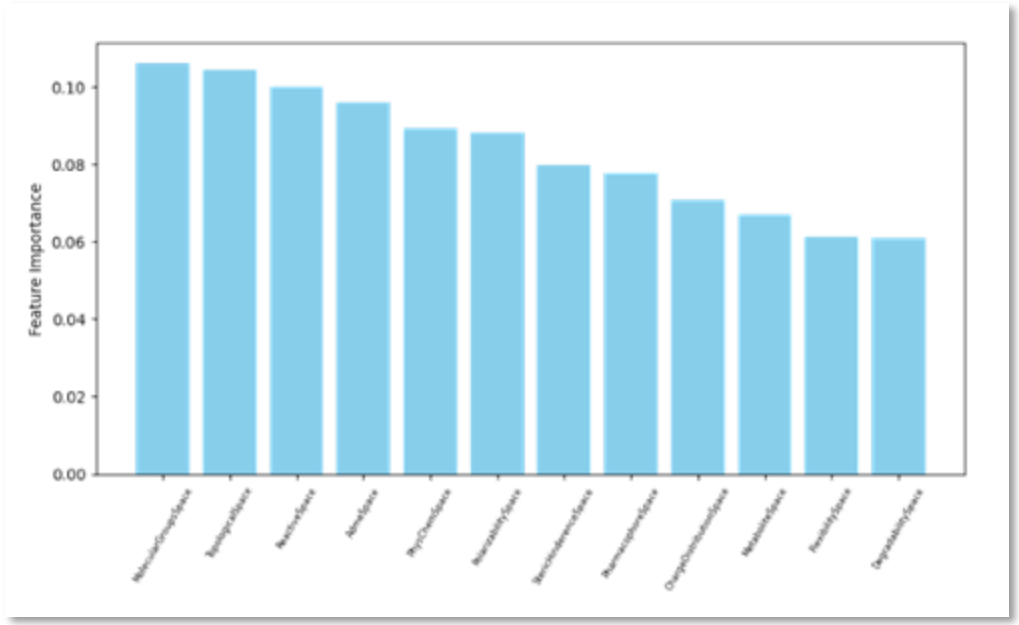


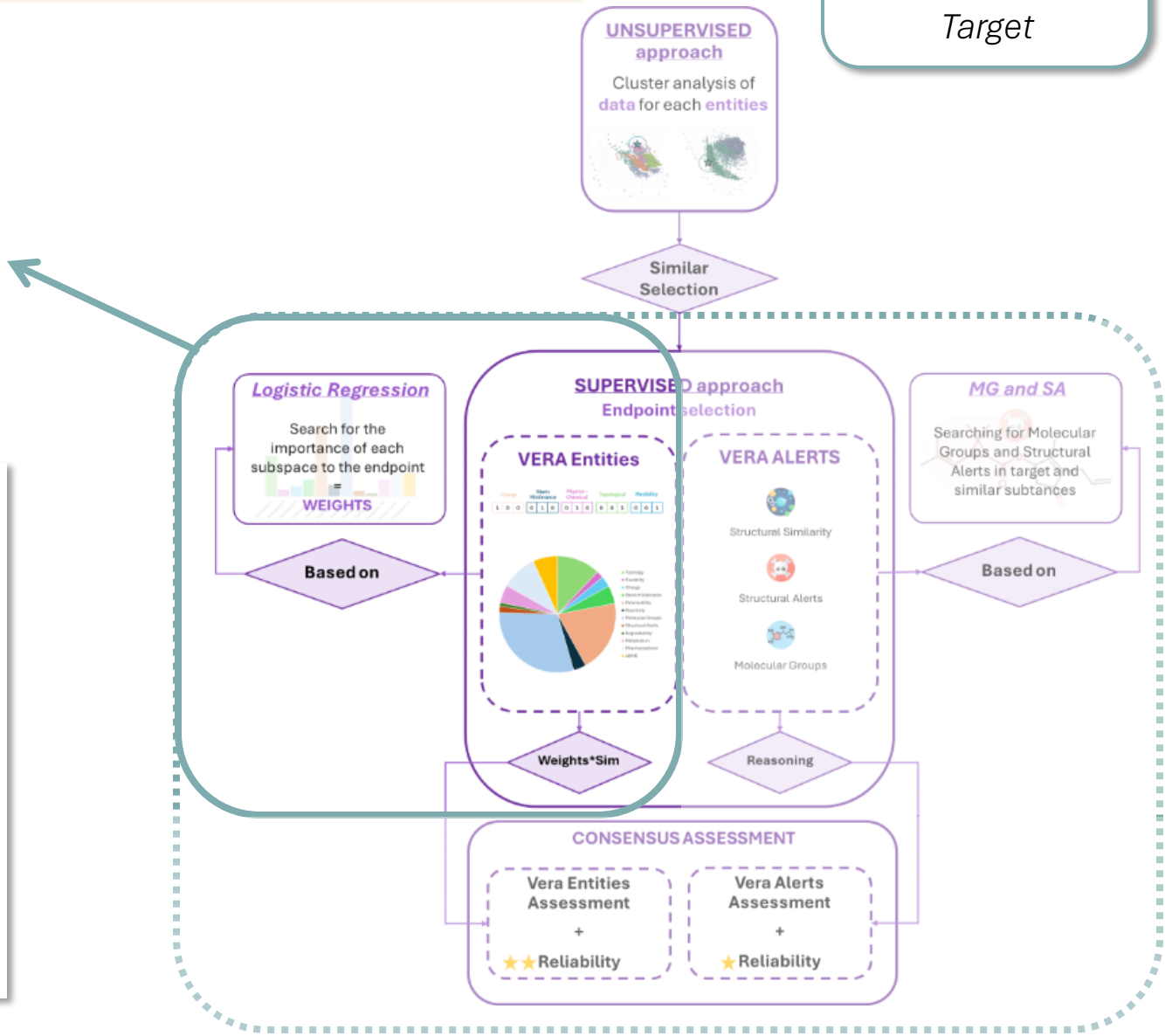
Virtual Extensive Read-Across: VERA



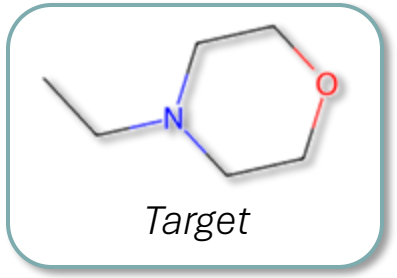
Goal: **endpoint assessment**

- VERA Entities
  - Entities importance
  - Similar (with value) selection
  - Prediction



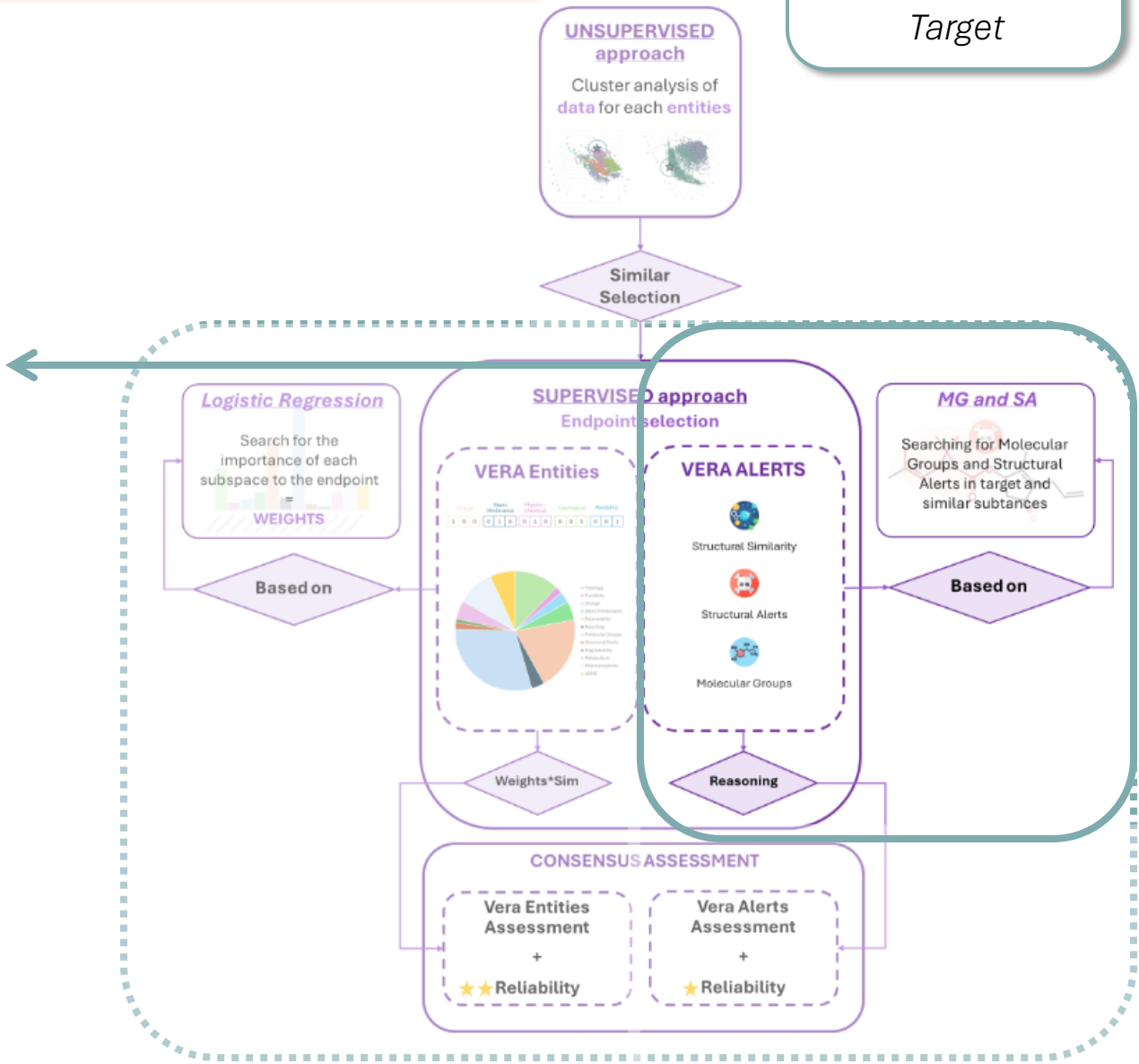


Virtual Extensive Read-Across: VERA

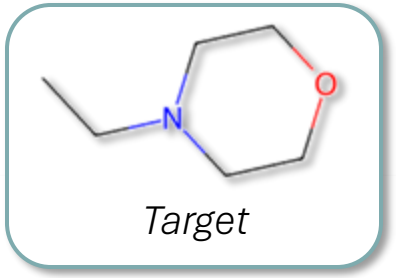


Goal: **endpoint assessment**

- VERA Entities
- VERA Alerts
  - Selection 1: Entities
  - Selection 2: Structural Alerts and Molecular Groups
  - Reasoning



Virtual Extensive Read-Across: VERA



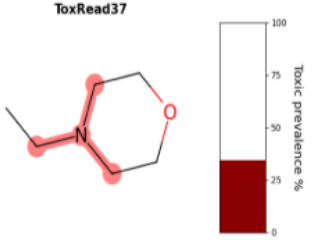
Target

VERA Alerts: similar substances used for the assessment

Endpoint: Fish Acute Toxicity LC50 (mg/L)

SMILES: CCN1CCOCC1

ToxRead37

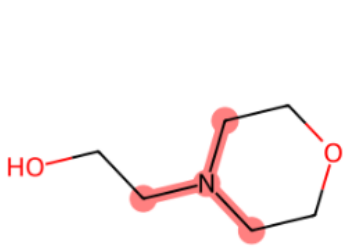


Warning:

Expert-based structural alerts (SAs) were found in the target molecule:

ToxRead37

Fish acute toxicity alert ToxRead37



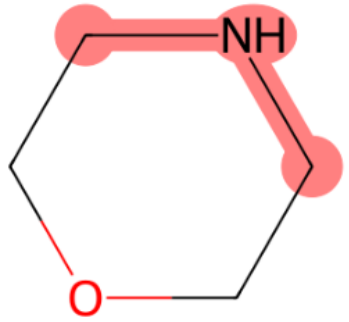
SMILES:

OCCN1CCOCC1

Experimental value mg/l:

2716.439

- VEGA Similarity: 0.906



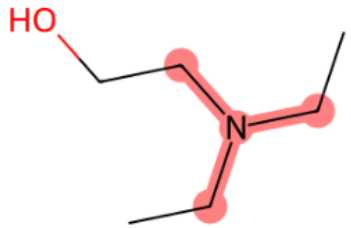
SMILES:

C1COCCN1

Experimental value mg/l:

181.97

- VEGA Similarity: 0.879



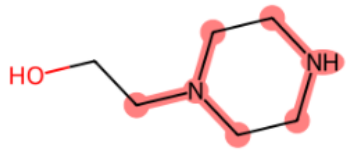
SMILES:

CCN(CC)CCO

Experimental value mg/l:

1774.189

- VEGA Similarity: 0.854



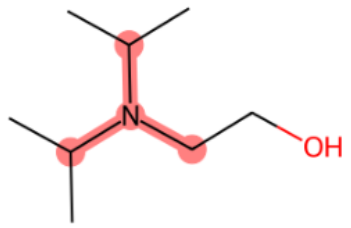
SMILES:

OCCN1CCNCC1

Experimental value mg/l:

6382.635

- VEGA Similarity: 0.806



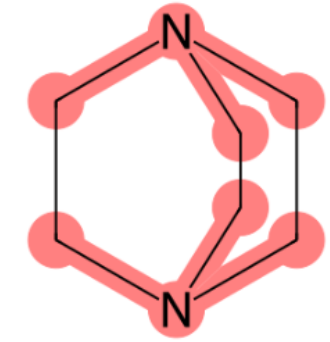
SMILES:

CC(C)N(CCO)C(C)C

Experimental value mg/l:

200.447

- VEGA Similarity: 0.799



SMILES:

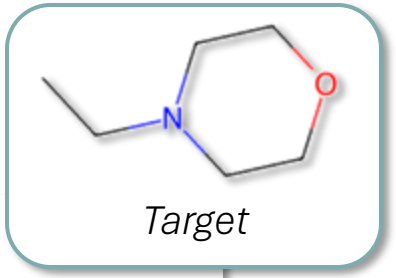
C1CN2CCN1CC2

Experimental value mg/l:

1729.816

- VEGA Similarity: 0.792

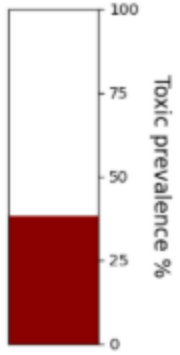
Virtual Extensive Read-Across: VERA



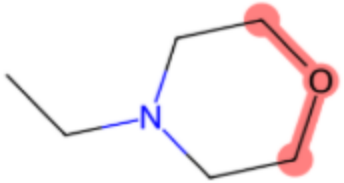
VERA Details

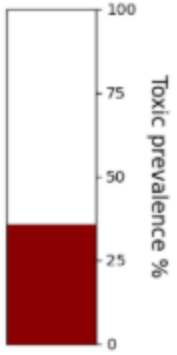
Molecular Groups

NH0

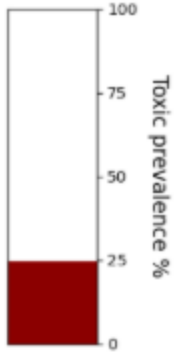


ether

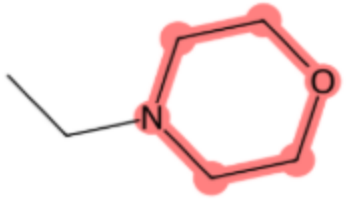


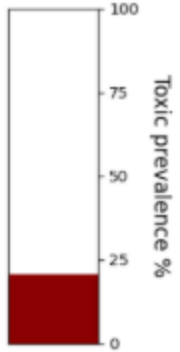


Ndealkylation1

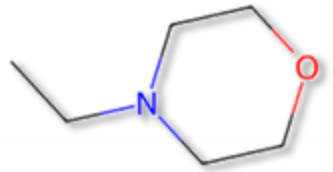


morpholine





Virtual Extensive Read-Across: VERA

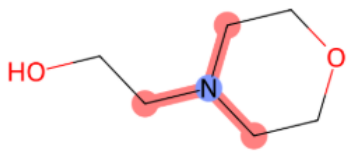


Reasoning

Coopresence analysis of:

1. Toxicity substructure (in red): could be expert base substructure or statistical relevant substructure
2. Specific molecular groups found in target [MGs]

MG: NH0



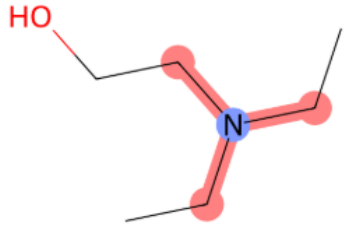
SMILES:

OCCN1CCOCC1

Experimental value (mg/l):

2716.439

- VEGA Similarity: 0.906



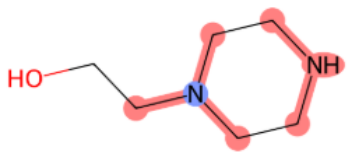
SMILES:

CCN(CC)CCO

Experimental value (mg/l):

1774.189

- VEGA Similarity: 0.854



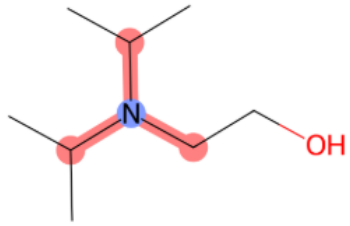
SMILES:

OCCN1CCNCC1

Experimental value (mg/l):

6382.635

- VEGA Similarity: 0.806



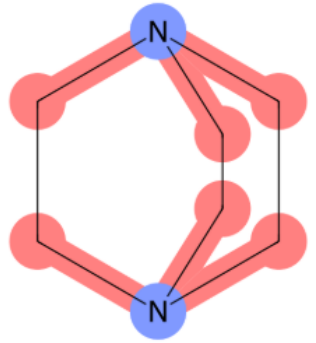
SMILES:

CC(C)N(CCO)C(C)C

Experimental value (mg/l):

200.447

- VEGA Similarity: 0.799



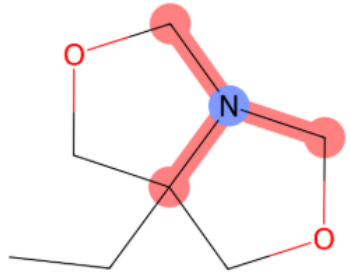
SMILES:

C1CN2CCN1CC2

Experimental value (mg/l):

1729.816

- VEGA Similarity: 0.792



SMILES:

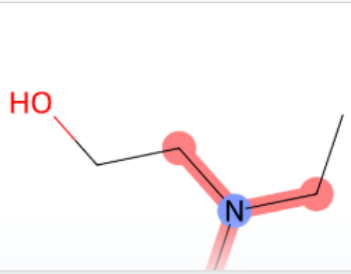
CCC12COCN1COC2

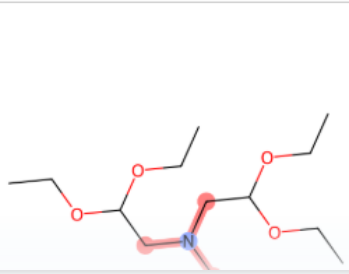
Experimental value (mg/l):

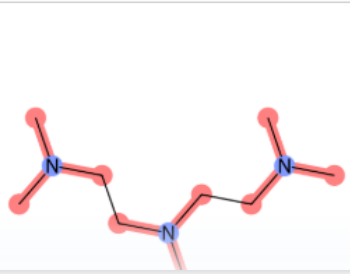
175.792

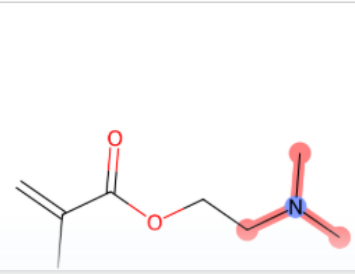
- VEGA Similarity: 0.781

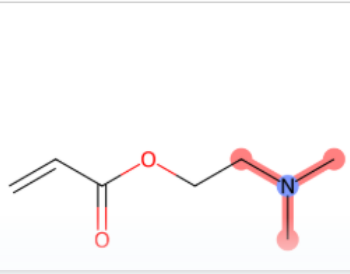
MG: Ndealkylation1



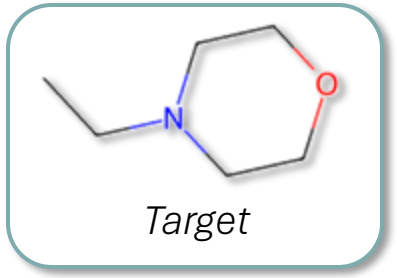






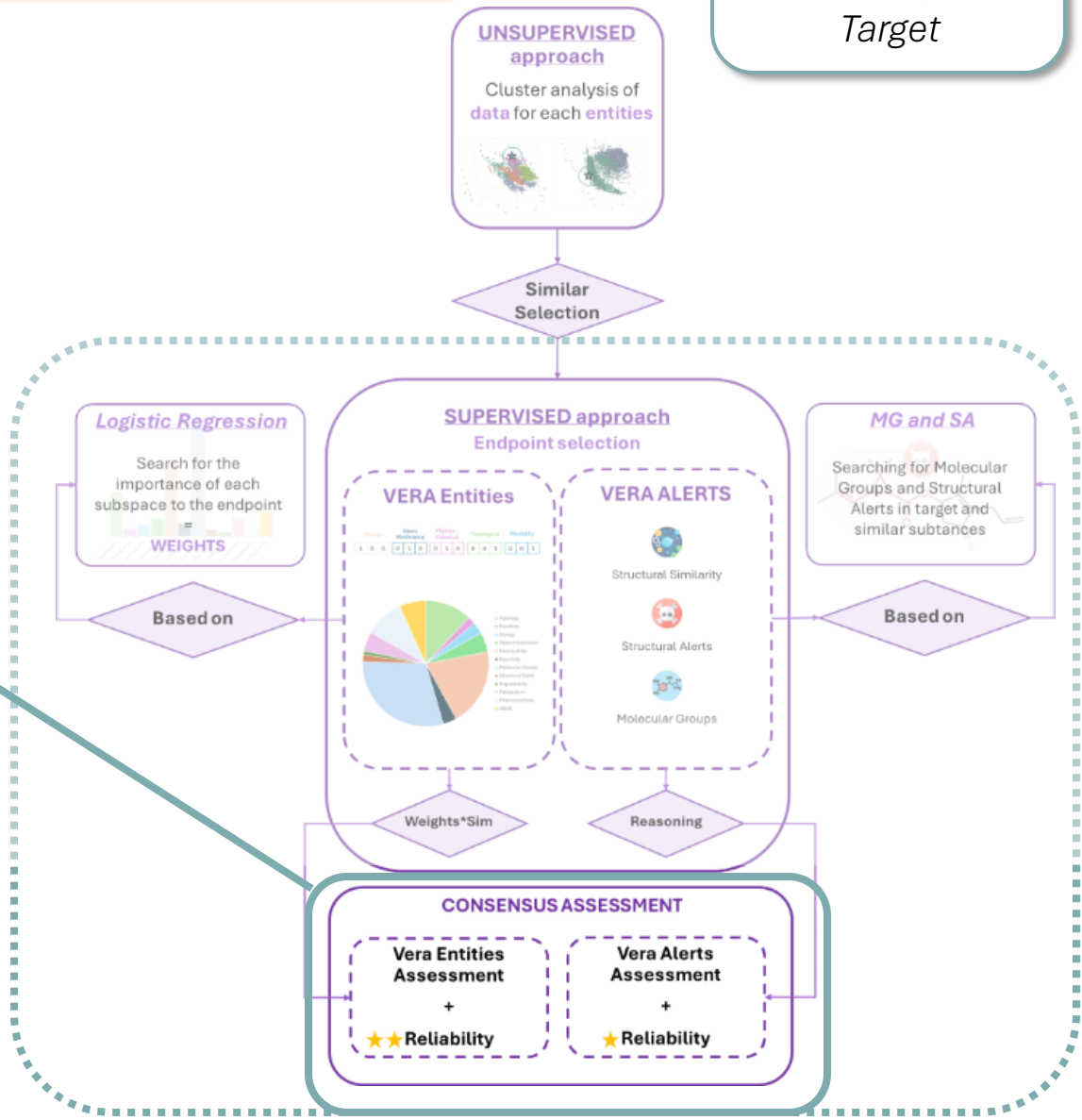


Virtual Extensive Read-Across: VERA



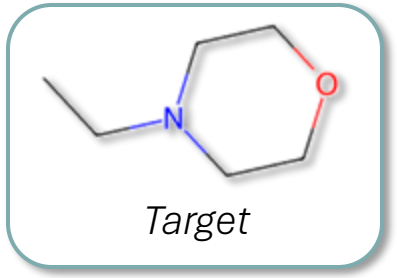
Goal: **endpoint assessment**

- VERA Entities
- VERA Alerts
- Consensus and **reliability**



Virtual Extensive Read-Across: VERA

UNSUPERVISED



Target

Assessment Summary

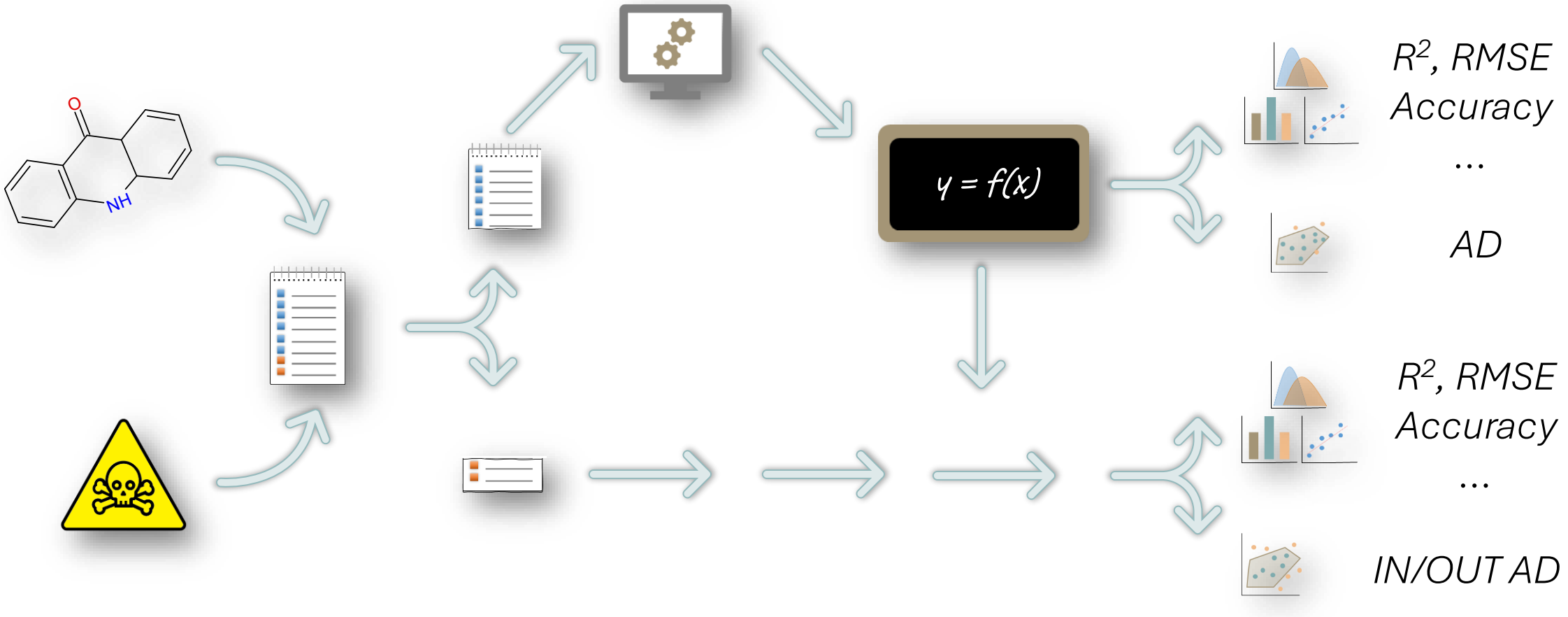
VERA Alerts (mg/l)	Reliability	VERA Entities (mg/l)	**Consensus**
606.644	High reliability	[223.872]	[415.258]

Chemicals used for Read Across

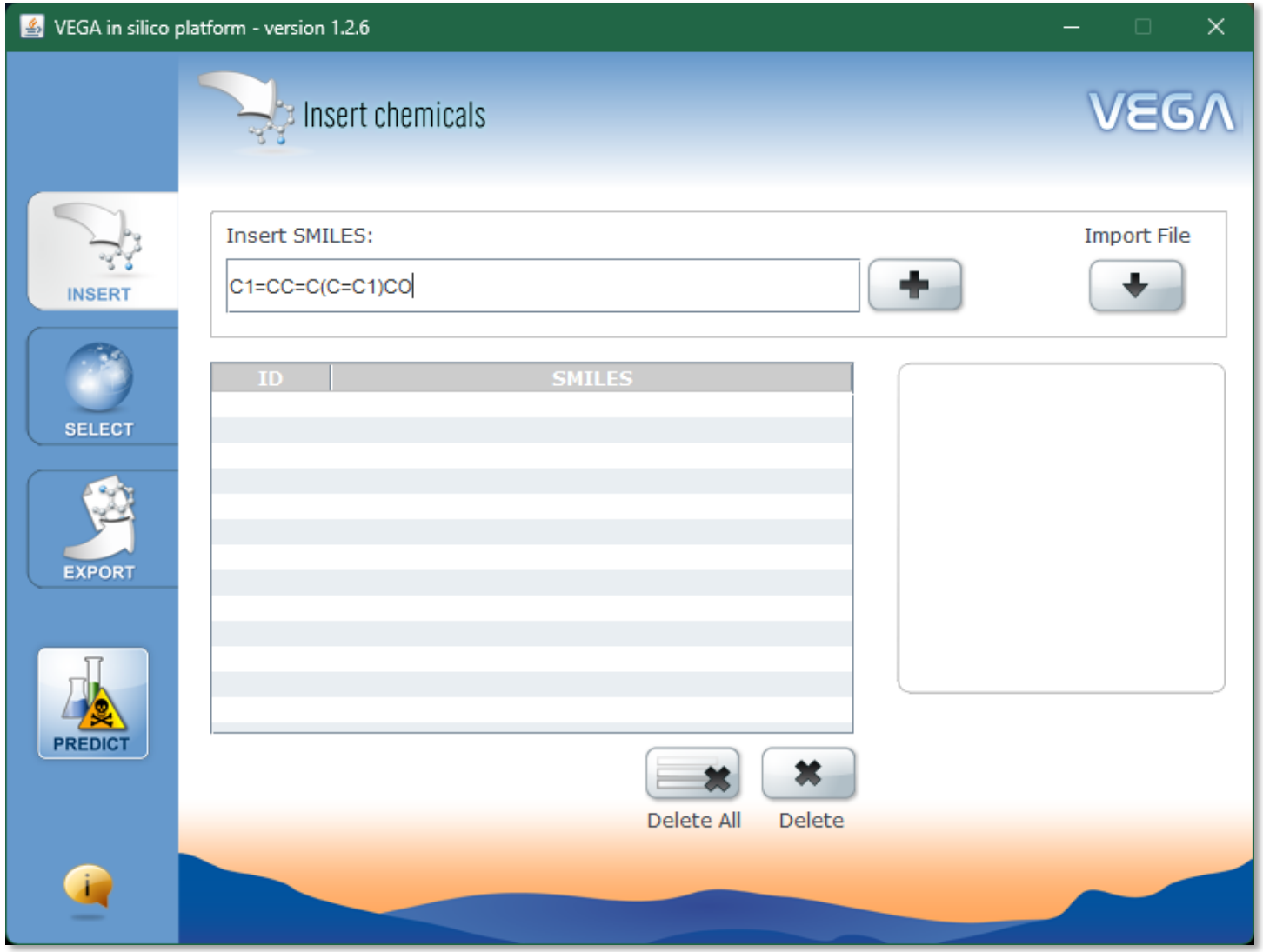
SMILES	VEGA_similarity	Experimental value mg/l	MGs	Alerts	Module
CCN1CCOCC1			NH0, Ndealkylation1, ether, morpholine	ToxRead37-Fish acute toxicity alert ToxRead37,	Target
OCCN1CCOCC1	0.906	2716.439	Al_OH, HOCCN, NH0, ether, morpholine	ToxRead37-Fish acute toxicity alert ToxRead37,	A, E
C1COCCN1	0.879	181.97	NH1, ether, morpholine	ToxRead37-Fish acute toxicity alert ToxRead37,	A, E
CCN(CC)CCO	0.854	1774.189	Al_OH, NH0, Ndealkylation1	ToxRead37-Fish acute toxicity alert ToxRead37,	A
OCCN1CCNCC1	0.806	6382.635	Al_OH, HOCCN, NH0, NH1, piperzine	ToxRead37-Fish acute toxicity alert ToxRead37,	A, E
CC(C)N(CCO)C(C)C	0.799	200.447	Al_OH, NH0	ToxRead37-Fish acute toxicity alert ToxRead37,	A
C1CN2CCN1CC2	0.792	1729.816	NH0, Al_bicycle, piperzine	ToxRead37-Fish acute toxicity alert ToxRead37,	A
CCN(CC)CC	0.774	23.878	NH0	ToxRead37-Fish acute toxicity alert ToxRead37,	E
CCOC(CN(C)CC(OCC)OCC)OCC	0.734	632.412	NH0, Ndealkylation1, ether	ToxRead37-Fish acute toxicity alert ToxRead37,	E
CN(C)CCN(C)CCN(C)C	0.698	109.648	NH0, Ndealkylation1	ToxRead37-Fish acute toxicity alert ToxRead37,	E
CCOc1cc(N2CCOCC2)c(OCC)cc1N	0.586	21.928	ArN, NH0, aniline, ether, morpholine, aniline_term	ToxRead37-Fish acute toxicity alert ToxRead37, ToxRead39-Fish acute toxicity alert ToxRead39,	E

(Q)SAR (Quantitative) structure-activity relationship

➡ estimates a property/activity using an **equation**

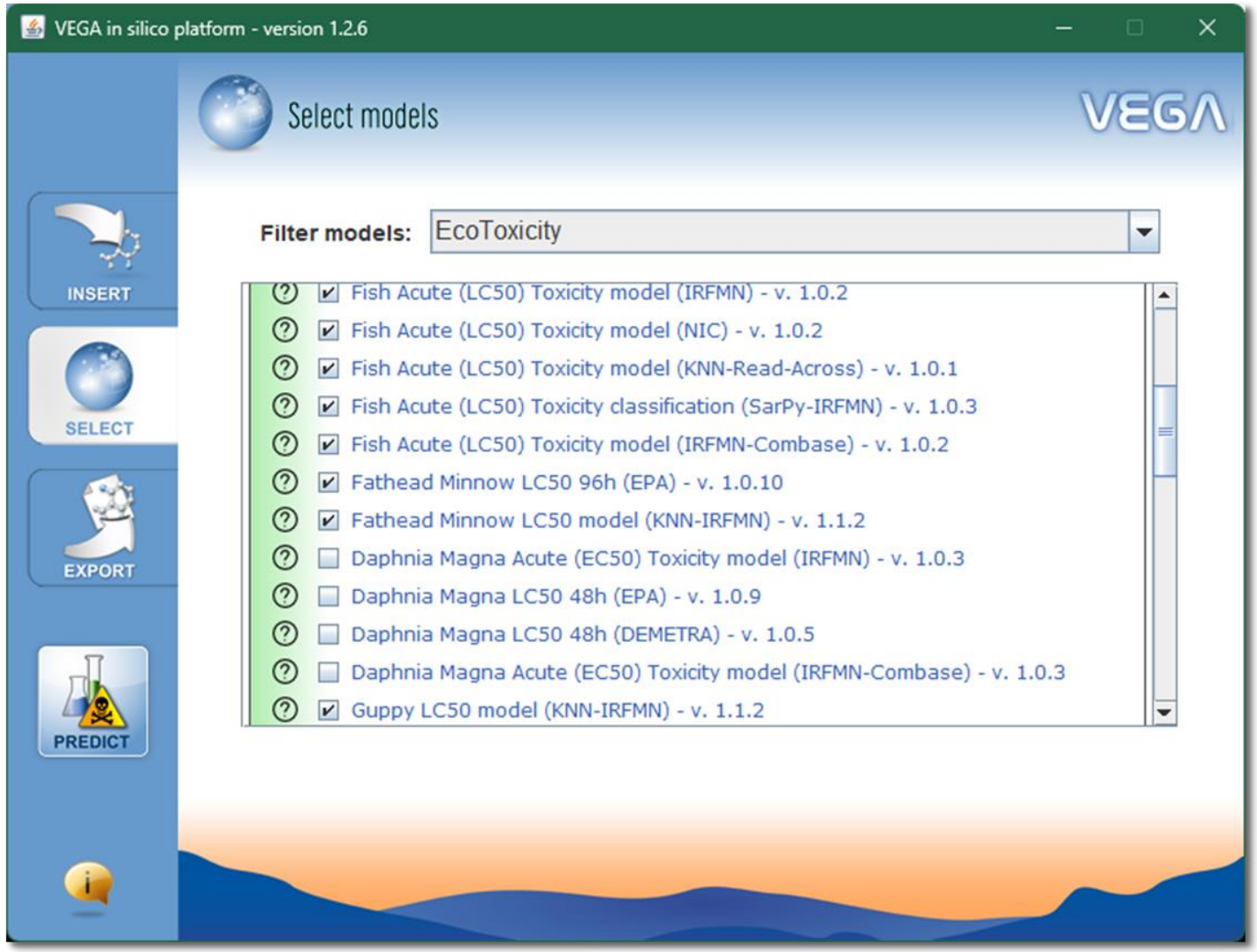


VEGA QSAR



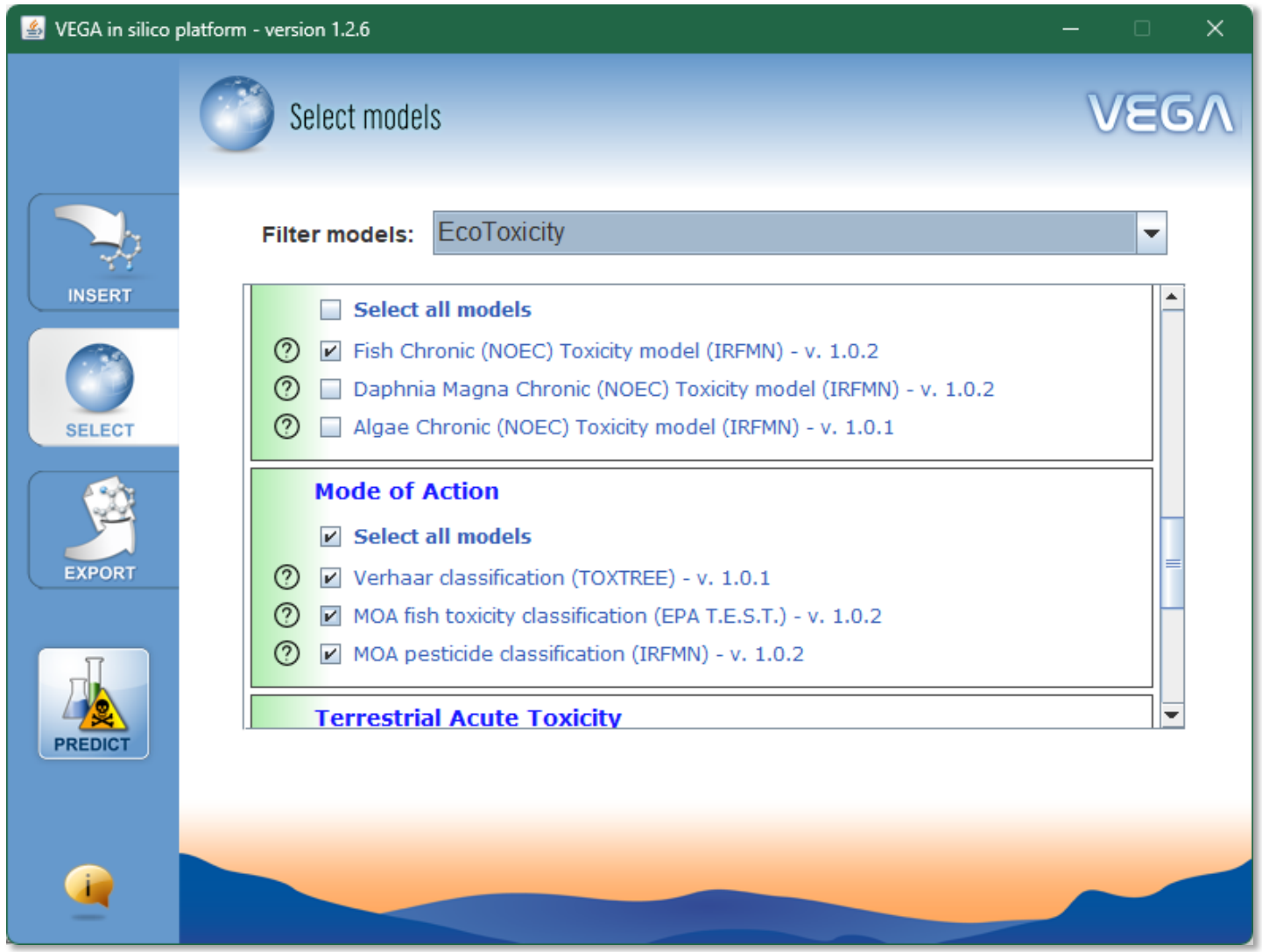
- 130+ models
- Toxicological, **ecotox**, environmental, phys-chem, toxicokinetic
- Applicability domain
- QMRF (https://www.vegahub.eu/portfolio-item/vega-qsar-models-qrmf/)

VEGA QSAR



- 5 **general models** for fish (1 in classification)
- 3 **species-specific** models

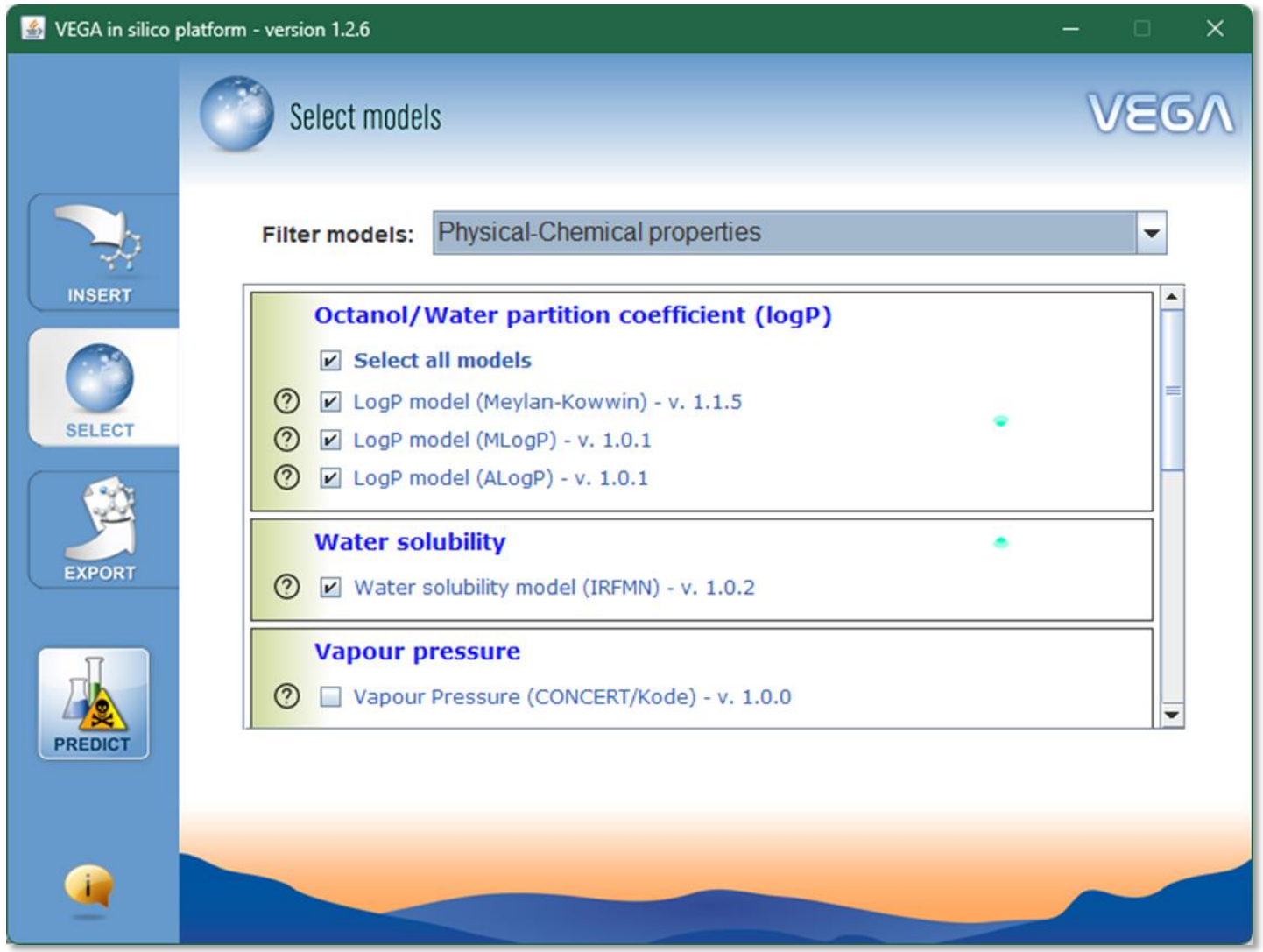
VEGA QSAR



Additional information

- Fish chronic toxicity
- MoA

VEGA QSAR



Additional information

- Fish chronic toxicity
- MoA
- Log Kow
- Water solubility

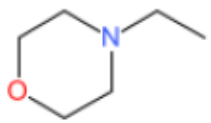
The PDF output

Summary page

- *Target*
- *Prediction*
- *Reliability*
- *Experimental value (if available)*
- *Eventual remarks*

1. Prediction Summary

Prediction for compound Molecule 1 -



Prediction: Reliability:

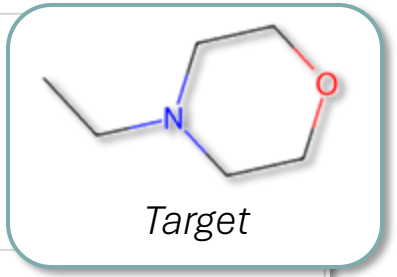
Prediction is 12.76 mg/L, but the result shows some critical aspects, which require to be checked:
- Only moderately similar compounds with known experimental value in the training set have been found

Compound: Molecule 1
Compound SMILES: O1CCN(CC)CC1
Experimental value: -
Predicted LC50 [a-dimensional]: -1.95
Predicted LC50 [mg/l]: 12.76
Molecular Weight: 115.2
Experimental value [mg/l]: -
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

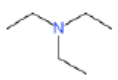
The PDF output

The six most similar chemicals

- *CAS and SMILES*
- *Training or test set*
- *VEGA similarity*
- *Experimental and predicted values*

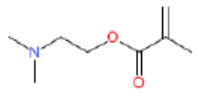


Target



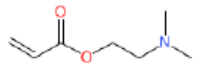
Compound #1

CAS: 121-44-8
Dataset id:335 (Training Set)
SMILES: N(CC)(CC)CC
Similarity: 0.774
Experimental value : -1.33
Predicted value : -1.331



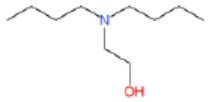
Compound #2

CAS: 2867-47-2
Dataset id:542 (Training Set)
SMILES: O=C(OCCN(C)C)C(=C)C
Similarity: 0.769
Experimental value : -1.885
Predicted value : -1.936



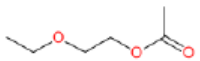
Compound #3

CAS: 2439-35-2
Dataset id:536 (Training Set)
SMILES: O=C(OCCN(C)C)C=C
Similarity: 0.766
Experimental value : -2.427
Predicted value : -2.204



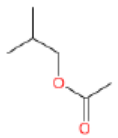
Compound #4

CAS: 102-81-8
Dataset id:215 (Training Set)
SMILES: OCCN(CCCC)CCCC
Similarity: 0.747
Experimental value : -1.622
Predicted value : -1.679



Compound #5

CAS: 111-15-9
Dataset id:288 (Training Set)
SMILES: O=C(OCCOCC)C
Similarity: 0.741
Experimental value : -1.076
Predicted value : -1.436



Compound #6

CAS: 110-19-0
Dataset id:271 (Training Set)
SMILES: O=C(OCC(C)C)C
Similarity: 0.737
Experimental value : -1.732
Predicted value : -1.935

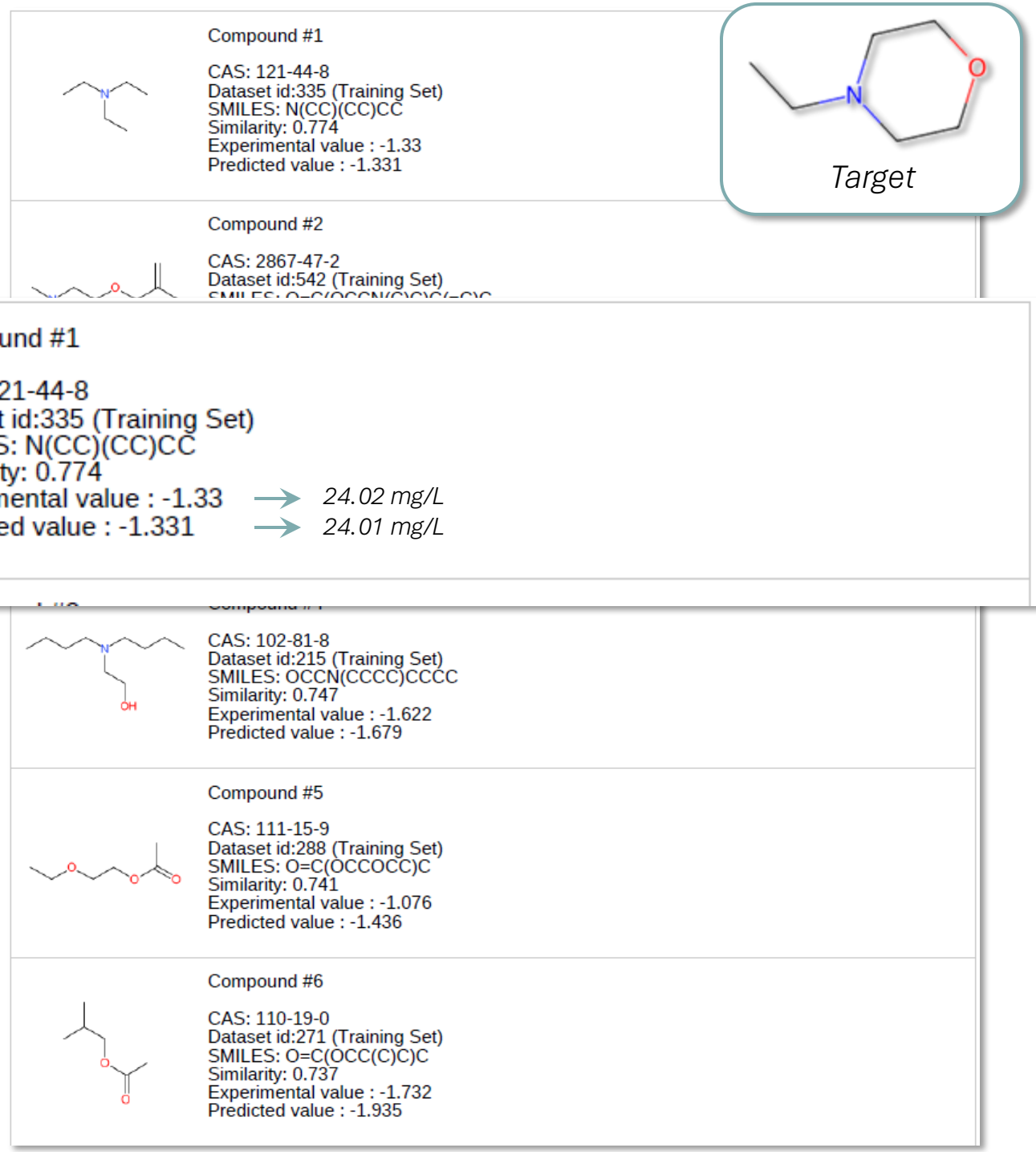
The PDF output

The six most similar chemicals

- *CAS and SMILES*
- *Training or test set*
- *VEGA similarity*
- *Experimental and predicted values*

Pay attention to the measure unit!!!

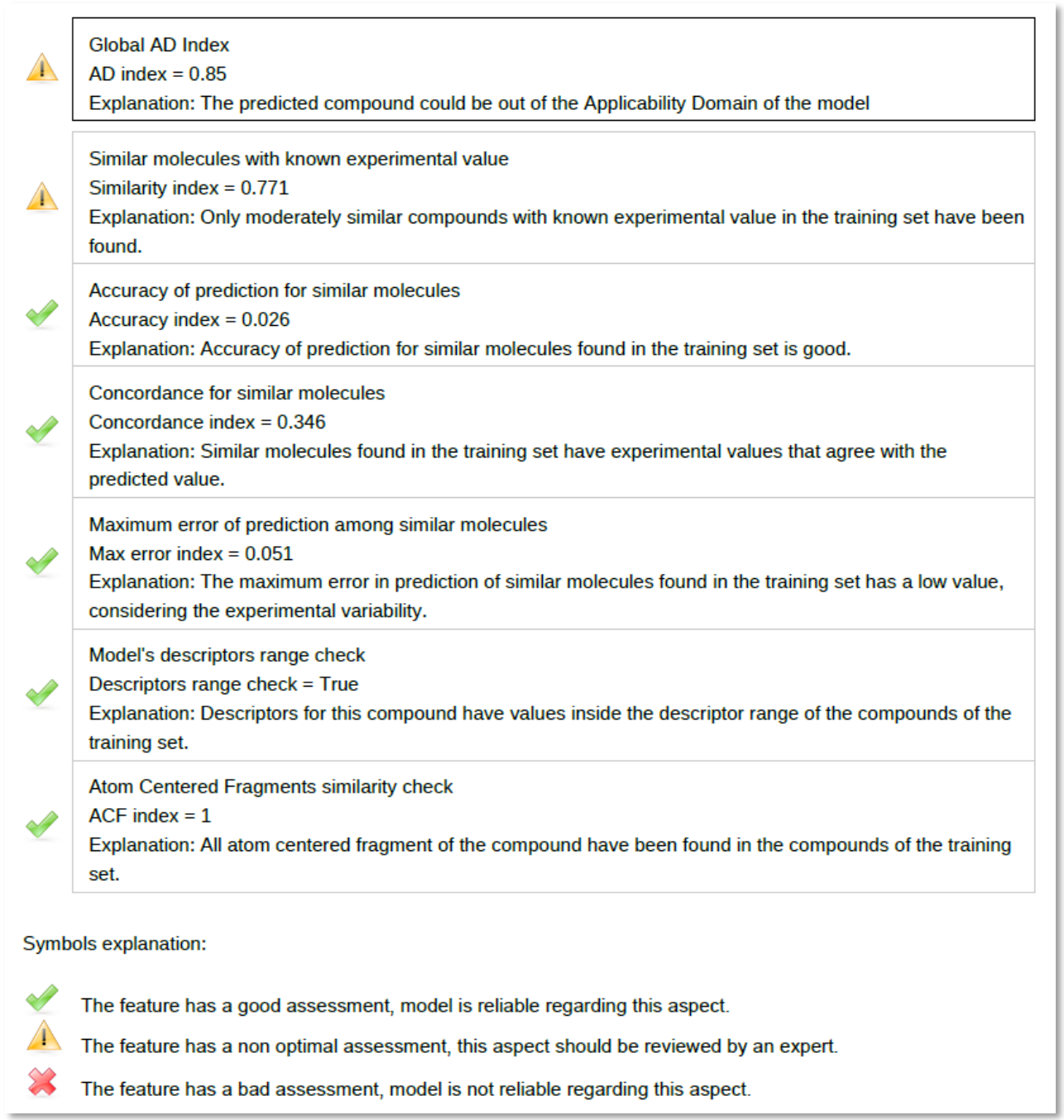
$$LD50_{mg/L} = \left((LD50_{a-dim} * 0.11 + 1)^{(1.0/0.11)}\right) * MW$$



The PDF output

The Applicability Domain Index (ADI)

- *Global AD index [0-1]*



Global AD Index
AD index = 0.85
Explanation: The predicted compound could be out of the Applicability Domain of the model

Similar molecules with known experimental value
Similarity index = 0.771
Explanation: Only moderately similar compounds with known experimental value in the training set have been found.

Accuracy of prediction for similar molecules
Accuracy index = 0.026
Explanation: Accuracy of prediction for similar molecules found in the training set is good.

Concordance for similar molecules
Concordance index = 0.346
Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.

Maximum error of prediction among similar molecules
Max error index = 0.051
Explanation: The maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability.

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:

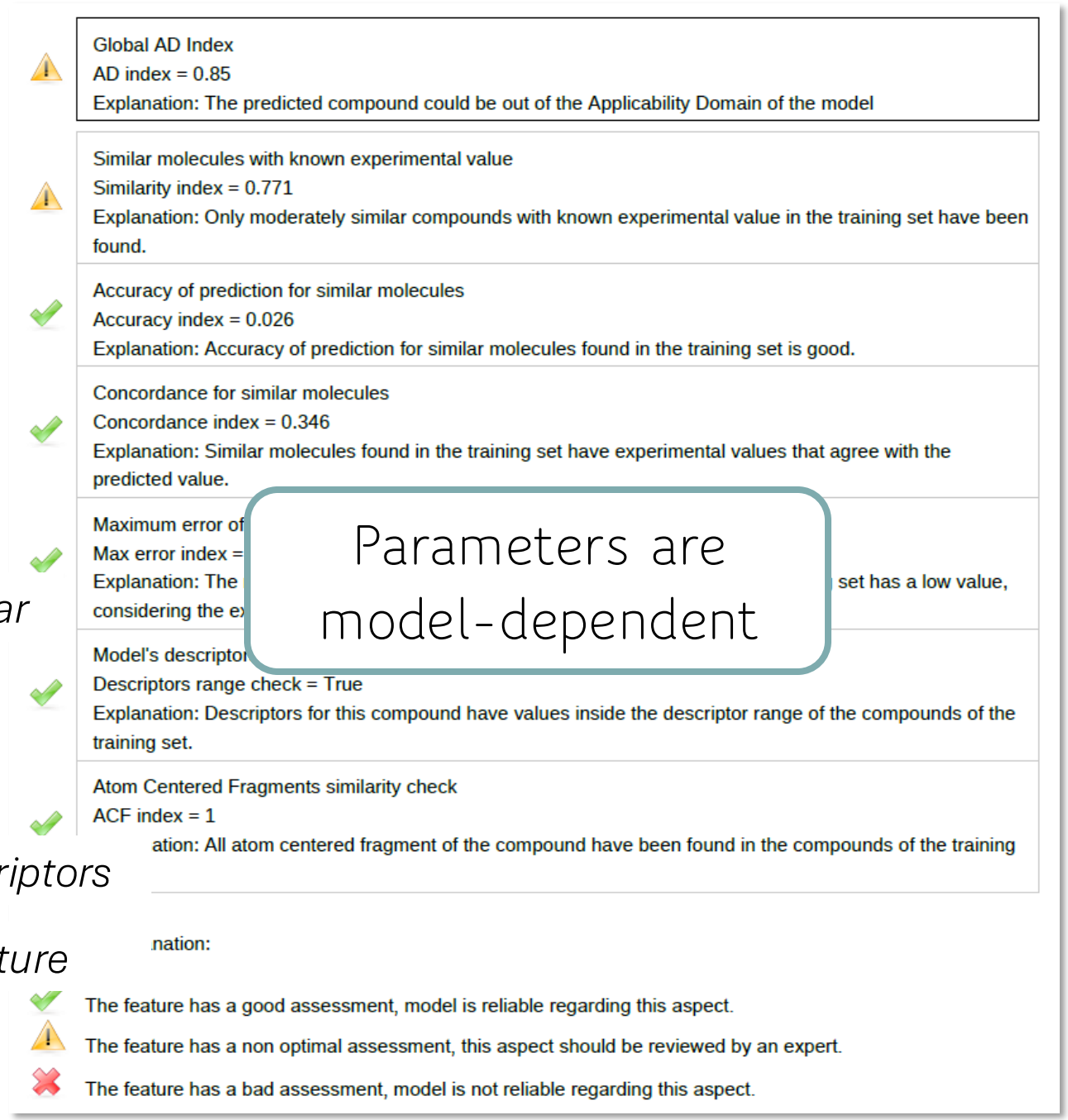
The feature has a good assessment, model is reliable regarding this aspect.
The feature has a non optimal assessment, this aspect should be reviewed by an expert.
The feature has a bad assessment, model is not reliable regarding this aspect.

The PDF output

The Applicability Domain Index (ADI)

- *Global AD index [0-1]*
  - *Similarity*
  - *Accuracy for similar*
  - *Concordance (target prediction – similar experimental)*
  - *Maximum error in prediction*

  Similar
  - *Descriptors range* — *Descriptors*
  - *Atom Centered Fragments* — *Structure*



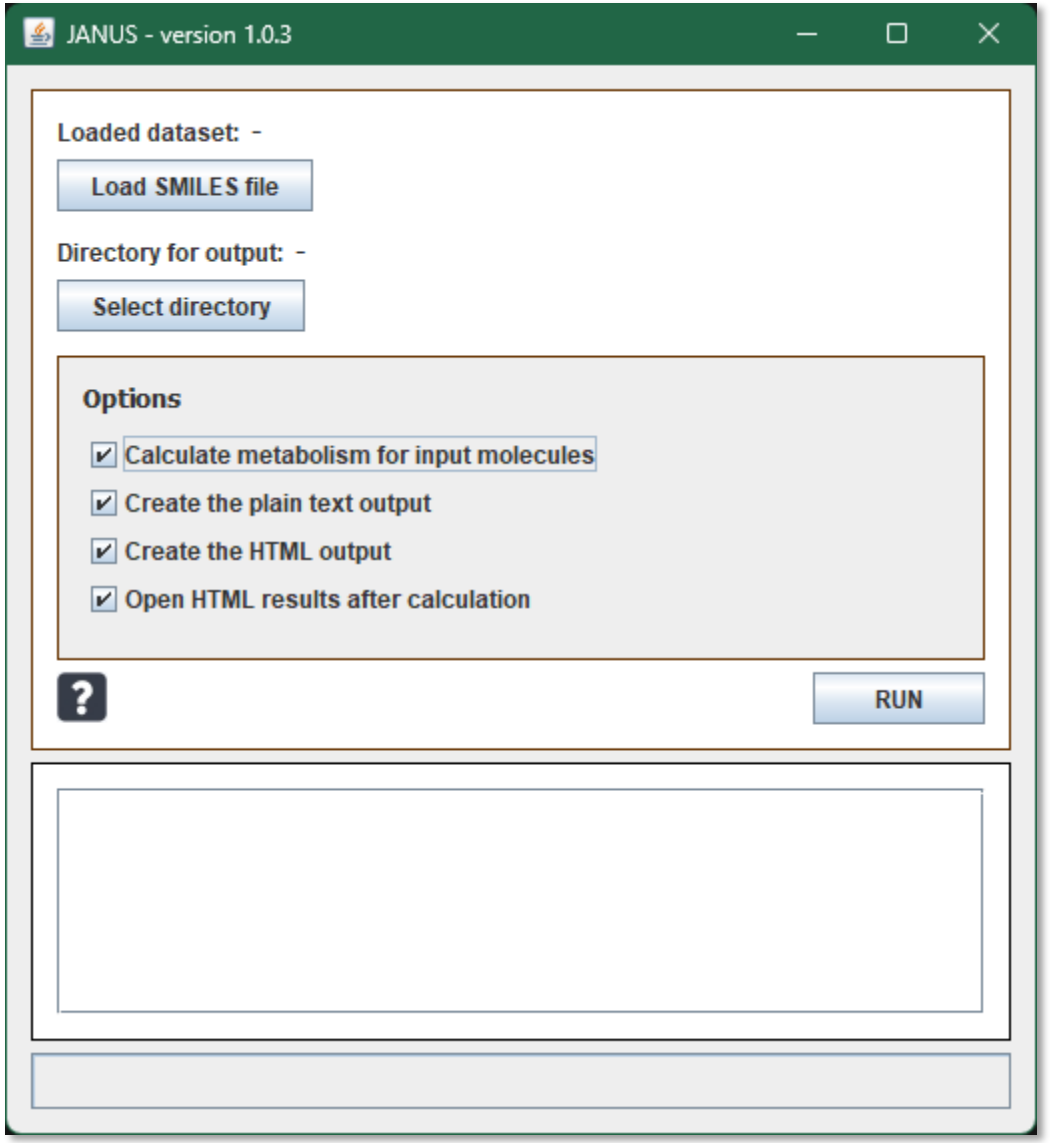
Parameters are model-dependent

Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit

Umwelt Bundesamt

The JANUS prioritization tool

Prioritization of chemical substances for **PBT**, **CMR** and **endocrine disrupting** activity



ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

kode from data to knowledge

KnowledgeMiner Software

JANUS vs VEGA QSAR

- Prioritization
- Consensus
- Additional parameters (e.g. water solubility)
- Microbial metabolites*
- VEGA QSAR output more detailed

**Microbial biocatalytic reactions and biodegradation pathways from EAWAG Biocatalysis/Biodegradation Database (https://envipath.org).*

The HTML output

Janus Result

NUMBER OF COMPOUNDS **1**

PBT CMR ED PARTIAL SCORES FINAL SCORES

	No.	Metabolite	Id	SMILES	Label	P	rel.	score	B [log(L/kg)]	rel.	score	T [mg/l]	rel.	score	C	rel.	score	M	rel.	score	R	rel.	score	ED	rel.	score	Score(vPvB)	Score(SVHC)	Score(PBT)
	1		Molecule 1	O1CCN(CC)CC1	P B T - C M R E	**nP**	0.59	0.289	**0.05**	0.99	0.051	**5.77**	0.51	0.249	**NC**	0.8	0.142	**M**	1	1	**T**	0.6	0.887	**NA**	0.8	0.142	0.121	0.909	0.237

Export as plain text

Molecule ID: Molecule 1

Back to index

Molecule

Info

Molecule number: 1
Molecule id: Molecule 1
Molecule SMILES: O1CCN(CC)CC1

Note

SUMMARY

Persistence **nP** (reliability: 0.59)	Bioaccumulation **0.05** (reliability: 0.99)	Toxicity **5.77** (reliability: 0.51)
Carcinogenicity **NON Carcinogenic** (reliability: 0.8)	Mutagenicity **MUTAGENIC** (reliability: 1)	Reproductive toxicity **TOXICANT** (reliability: 0.6)
Endocrine Disruptor **Inactive** (reliability: 0.8)		

Toxicity assessment

Overall prediction [mg/l]	Overall Reliability
5.77	**0.51**

Values retrieved and/or calculated in the workflow:

Property	Value
Fish Acute (LC50) Toxicity model (KNN)	1039.26 mg/L (good reliability)
Fish Acute (LC50) Toxicity model (NIC)	155.59 mg/L (moderate reliability)
Fish Acute (LC50) Toxicity model (IRFMN)	12.7 mg/L (moderate reliability)
Fathead Minnow Acute (LC50) Toxicity model (EPA)	1831.3 mg/L (good reliability)
Fathead Minnow Acute (LC50) Toxicity model (KNN)	1083.88 mg/L (moderate reliability)
Guppy Acute (LC50) Toxicity model (KNN)	N/A
Fish Chronic (NOEC) Toxicity model (IRFMN)	5.77 mg/L (low reliability)
Daphnia Magna Acute (LC50) Toxicity model (DEMETRA)	12.31 mg/L (moderate reliability)
Daphnia Magna Acute (LC50) Toxicity model (EPA)	44.53 mg/L (low reliability)
Daphnia Magna Acute (LC50) Toxicity model (IRFMN)	195.47 mg/L (moderate reliability)
Daphnia Magna Chronic (NOEC) Toxicity model (IRFMN)	5.93 mg/L (moderate reliability)
Algae Acute (EC50) Toxicity model (IRFMN)	25.39 mg/L (moderate reliability)
Algae Chronic (NOEC) Toxicity model (IRFMN)	16.55 mg/L (low reliability)
Water Solubility model (IRFMN)	995697.81 mg/L (EXPERIMENTAL value)
Overall Fish Acute (LC50) Toxicity assessment (used in the workflow)	1379.6 mg/L (reliability: 0.83)
Overall Fish Chronic (NOEC) Toxicity assessment (used in the workflow)	5.77 mg/L (reliability: 0.45)
Overall Fish assessment (used in the workflow)	NON Toxic, 5.77 mg/L (reliability: 0.45)
Overall Daphnia Magna Acute (LC50) Toxicity assessment (used in the workflow)	103.9 mg/L (reliability: 0.68)
Overall Daphnia Magna Chronic (NOEC) Toxicity assessment (used in the workflow)	5.93 mg/L (reliability: 0.65)
Overall Daphnia Magna assessment (used in the workflow)	NON Toxic, 5.93 mg/L (reliability: 0.65)
Overall Algae Acute (EC50) Toxicity assessment (used in the workflow)	25.4 mg/L (reliability: 0.65)
Overall Algae Chronic (NOEC) Toxicity assessment (used in the workflow)	16.6 mg/L (reliability: 0.45)
Overall Algae assessment (used in the workflow)	NON Toxic, 16.6 mg/L (reliability: 0.45)

Janus Result

NUMBER OF COMPOUN

FINAL SCORES

	No.	Metaboli	VHC)	Score(PBT)
	1			0.237

Export as plain text

Molecule ID: Molecule 1

Back to index

Molecule

Info

Molecule number: 1
Molecule id: Molecule 1
Molecule SMILES: O1CCN(CC)CC1

Note

SUMMARY

Persistence
nP (reliability: 0.59)

Bioaccumulation
0.05 (reliability: 0.99)

Toxicity
5.77 (reliability: 0.51)

Carcinogenicity
NON Carcinogenic (reliability: 0.8)

Mutagenicity
MUTAGENIC (reliability: 1)

Reproductive toxicity
TOXICANT (reliability: 0.6)

Janus Result

NUMBER OF COMPOUN

FINAL SCORES

	No.	Metaboli	VHC)	Score(PBT)
	1			0.237

Toxicity assessment

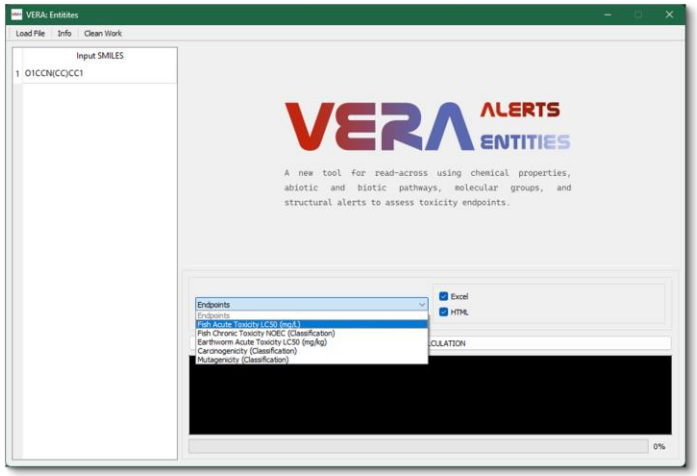
Values retrieved and/or calculated in the workflow:

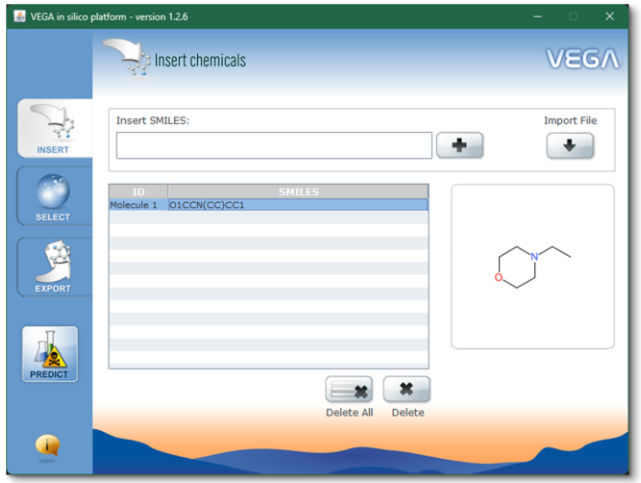
Property	Value
Fish Acute (LC50) Toxicity model (KNN)	1039.26 mg/L (good reliability)
Fish Acute (LC50) Toxicity model (NIC)	155.59 mg/L (moderate reliability)
Fish Acute (LC50) Toxicity model (IRFMN)	12.7 mg/L (moderate reliability)
Fathead Minnow Acute (LC50) Toxicity model (EPA)	1831.3 mg/L (good reliability)
Fathead Minnow Acute (LC50) Toxicity model (KNN)	1083.88 mg/L (moderate reliability)
Guppy Acute (LC50) Toxicity model (KNN)	N/A
Water Solubility model (IRFMN)	995697.81 mg/L (EXPERIMENTAL value)
Overall Fish Acute (LC50) Toxicity assessment (used in the workflow)	1379.6 mg/L (reliability: 0.83)
Algae Acute (EC50) Toxicity model (IRFMN)	25.39 mg/L (moderate reliability)
Algae Chronic (NOEC) Toxicity model (IRFMN)	16.55 mg/L (low reliability)
Overall Daphnia Magna Chronic (NOEC) Toxicity assessment (used in the workflow)	5.93 mg/L (reliability: 0.65)
Overall Daphnia Magna assessment (used in the workflow)	NON Toxic, 5.93 mg/L (reliability: 0.65)
Overall Algae Acute (EC50) Toxicity assessment (used in the workflow)	25.4 mg/L (reliability: 0.65)
Overall Algae Chronic (NOEC) Toxicity assessment (used in the workflow)	16.6 mg/L (reliability: 0.45)
Overall Algae assessment (used in the workflow)	NON Toxic, 16.6 mg/L (reliability: 0.45)

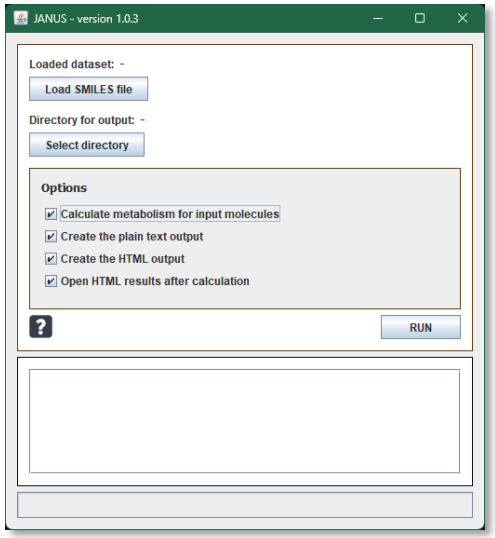
Export as plain text

Take-home message

Several ***in silico*** methods can be used to assess **fish acute toxicity**







Automatic read-across

Multiple approaches for similar selection

Reliability

QSAR models

ADI

Similar chemicals

Prioritization

Consensus (VEGA models)

Microbial metabolites

Next future: **SESAPI** (for pharmaceuticals)

Thank you for your attention!

anna.lombardo@marionegri.it

https://www.vegahub.eu/

ISTITUTO DI RICERCHE
FARMACOLOGICHE
MARIO NEGRI · IRCCS