VEGA QSAR MODELS

Report

Prediction and Applicability Domain analysis for models:

- Mutagenicity (Ames test) CONSENSUS model 1.0.4
- Mutagenicity (Ames test) model (CAESAR) 2.1.14
- Mutagenicity (Ames test) model (ISS) 1.0.3
- Mutagenicity (Ames test) model (SarPy-IRFMN) 1.0.8
- Mutagenicity (Ames test) model (KNN-Read-Across) 1.0.1
- Mutagenicity (Ames test) model for aromatic amines (CONCERT/IRFMN) 1.0.0
- Carcinogenicity model (CAESAR) 2.1.10
- Carcinogenicity model (ISS) 1.0.3
- Carcinogenicity model (IRFMN-ISSCAN-CGX) 1.0.2
- Carcinogenicity model (IRFMN-Antares) 1.0.2
- Carcinogenicity oral classification model (IRFMN) 1.0.1
- Carcinogenicity oral Slope Factor model (IRFMN) 1.0.1
- Carcinogenicity inhalation classification model (IRFMN) 1.0.1
- Carcinogenicity inhalation Slope Factor model (IRFMN) 1.0.1
- Carcinogenicity in male rat (CORAL) 1.0.0
- Carcinogenicity in female Rat (CORAL) 1.0.0
- Acute Toxicity (LD50) model (KNN) 1.0.0
- BCF model (CAESAR) 2.1.15
- BCF model (Meylan) 1.0.4
- BCF model (Arnot-Gobas) 1.0.1
- BCF model (KNN-Read-Across) 1.1.1

Core version: 1.3.18

Prediction for compound Molecule 1 -

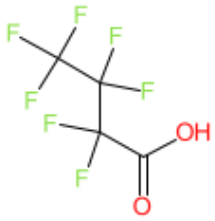
Prediction:

Prediction is NON-Mutagenic with a consensus score of 0.5, based on 4 models.

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Used models: 4
Predicted Consensus Mutagen activity: NON-Mutagenic
Mutagenic Score: 0
Non-Mutagenic Score: 0.5
Model Caesar assessment: NON-Mutagenic (MODERATE reliability)
Model ISS assessment: NON-Mutagenic (LOW reliability)
Model SarPy assessment: NON-Mutagenic (MODERATE reliability)
Model KNN assessment: NON-Mutagenic (MODERATE reliability)
Remarks:
none

1. Prediction Summary

Prediction for compound Molecule 1 -

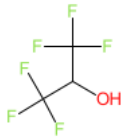


Prediction: Reliability:

Prediction is NON-Mutagenic, 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
- Accuracy of prediction for similar molecules found in the training set is not optimal

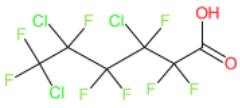
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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



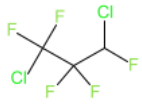
Compound #1

CAS: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #2

CAS: 2106-54-9
Dataset id:384 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl
Similarity: 0.831
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #3

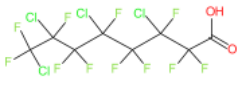
CAS: 507-55-1
Dataset id:34 (Training Set)
SMILES: FC(C(F)(F)C(F)(F)Cl)Cl
Similarity: 0.757
Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

Alerts (not found also in the target): SA8 Aliphatic halogens

Compound #4

CAS: 422-56-0
Dataset id:854 (Training Set)
SMILES: FC(F)(F)C(F)(F)C(Cl)Cl
Similarity: 0.752
Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

Alerts (not found also in the target): SA8 Aliphatic halogens



Compound #5

CAS: 2923-68-4
Dataset id:2832 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl)Cl
Similarity: 0.744
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #6

CAS: 335-76-2
Dataset id:3947 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.715
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.817 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.802 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.693 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✓	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.

1. Prediction Summary

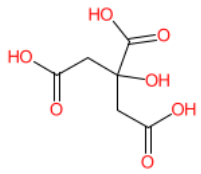
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is NON-Mutagenic, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)

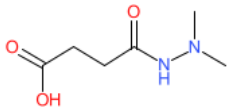
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



Compound #1

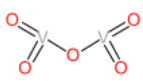
CAS: 77-92-9
Dataset id:829 (Training Set)
SMILES: O=C(O)CC(O)(C(=O)O)CC(=O)O
Similarity: 0.669
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #2

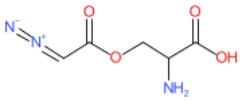
CAS: 1596-84-5
Dataset id:225 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.636
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA13 Hydrazine



Compound #3

CAS: 1314-62-1
Dataset id:771 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.632
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

CAS: 115-02-6
Dataset id:443 (Training Set)
SMILES: [N-]=[N+]=CC(=O)OCC(N)C(=O)O
Similarity: 0.629
Experimental value : Mutagenic
Predicted value : Mutagenic

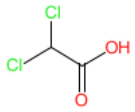
Alerts (not found also in the target): SA14 Aliphatic azo and azoxy

Compound #5

CAS: 75-88-7
Dataset id:819 (Training Set)
SMILES: FC(F)(F)CCl
Similarity: 0.628
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA8 Aliphatic halogens

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



Compound #6

CAS: 79-43-6
Dataset id:846 (Training Set)
SMILES: O=C(O)C(Cl)Cl
Similarity: 0.627
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA8 Aliphatic halogens

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.35 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.652 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.522 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✔	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.51 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)..

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.

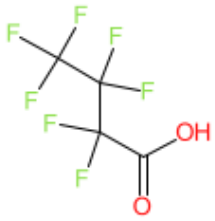
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CC(F)(F)F The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: CC(C)(F)F The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 1 -

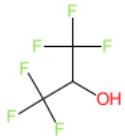


Prediction: Reliability:

Prediction is NON-Mutagenic, but the result shows some critical aspects, which require to be checked:
- Accuracy of prediction for similar molecules found in the training set is not adequate
The following relevant fragments have been found: SM150

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
No. alerts for mutagenicity: 0
No. alerts for non-mutagenicity: 1
Structural Alerts: SM150
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

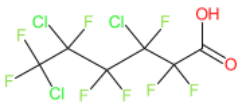
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM150

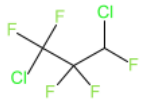


Compound #2

CAS: 2106-54-9
Dataset id:384 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl
Similarity: 0.831
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM106; SM177



Compound #3

CAS: 507-55-1
Dataset id:34 (Training Set)
SMILES: FC(C(F)(F)C(F)(F)Cl)Cl
Similarity: 0.757
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

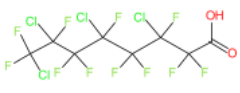
Alerts (not found also in the target): SM106

Compound #4

CAS: 422-56-0
Dataset id:854 (Training Set)
SMILES: FC(F)(F)C(F)(F)C(Cl)Cl
Similarity: 0.752
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM106



Compound #5

CAS: 2923-68-4
Dataset id:2832 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl)Cl
Similarity: 0.744
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM106; SM163; SM177

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

Compound #6

CAS: 335-76-2
Dataset id:3947 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.715
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM143; SM163; SM177

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.688 Explanation: The predicted compound could be out of the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.802 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✗	Accuracy of prediction for similar molecules Accuracy index = 0.347 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✓	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree 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..

Symbols explanation:

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

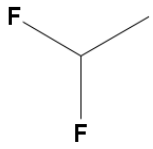
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

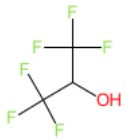
(Molecule 1) Reasoning on fragments/structural alerts :.

Fragment found: SM150



Sarpy alert n. 150 for NON-Mutagenicity, defined by SMARTS: C(F)(F)C

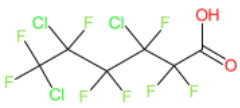
Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833

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

Alerts (found also in the target): SM150

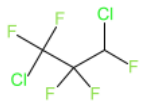


CAS: 2106-54-9
Dataset id:384 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl
Similarity: 0.831

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM106; SM177



CAS: 507-55-1
Dataset id:34 (Training Set)
SMILES: FC(C(F)(F)C(F)(F)Cl)Cl
Similarity: 0.757

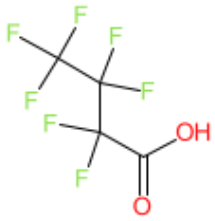
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM106

1. Prediction Summary

Prediction for compound Molecule 1 -

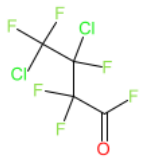


Prediction: Reliability:

Prediction is NON-Mutagenic, but the result shows some critical aspects, which require to be checked:
- Accuracy of prediction for similar molecules found in the training set is not adequate
- some similar molecules found in the training set have experimental values that disagree with the predicted value

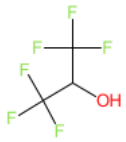
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Molecules used for prediction: 4
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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



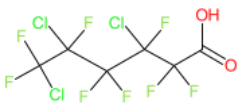
Compound #1

CAS: 678-06-8
Dataset id:5701 (Training Set)
SMILES: O=C(F)C(F)(F)C(F)(C(F)(F)Cl)Cl
Similarity: 0.862
Experimental value : Mutagenic
Predicted value : NON-Mutagenic



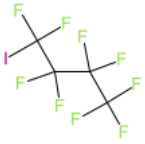
Compound #2

CAS: 920-66-1
Dataset id:5212 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #3

CAS: 2106-54-9
Dataset id:2157 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl
Similarity: 0.831
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

CAS: 423-39-2
Dataset id:3062 (Training Set)
SMILES: FC(F)(F)C(F)(F)C(F)(F)C(F)(F)I
Similarity: 0.804
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Compound #5

CAS: 13098-39-0
Dataset id:1107 (Training Set)
SMILES: O=C(C(F)(F)F)C(F)(F)F
Similarity: 0.803
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #6

CAS: 431-89-0
Dataset id:3090 (Training Set)
SMILES: FC(C(F)(F)F)C(F)(F)F
Similarity: 0.789
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.71 Explanation: The predicted compound could be out of the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.83 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✗	Accuracy of prediction for similar molecules Accuracy index = 0.5 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
⚠	Concordance for similar molecules Concordance index = 0.739 Explanation: some 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..

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.

1. Prediction Summary

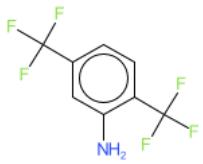
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is NA, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)

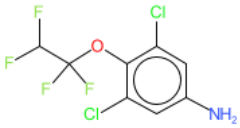
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NA
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



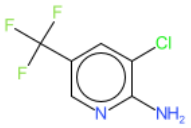
Compound #1

CAS: N.A.
Dataset id:7541 (Training Set)
SMILES: FC(F)(F)c1ccc(c(N)c1)C(F)(F)F
Similarity: 0.641
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



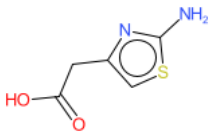
Compound #2

CAS: N.A.
Dataset id:5582 (Training Set)
SMILES: FC(F)C(F)(F)Oc1c(cc(N)cc1Cl)Cl
Similarity: 0.633
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



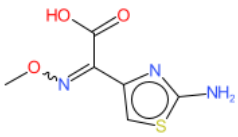
Compound #3

CAS: N.A.
Dataset id:4740 (Training Set)
SMILES: FC(F)(F)c1cnc(N)c(c1)Cl
Similarity: 0.603
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



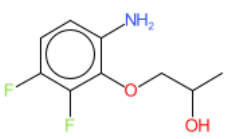
Compound #4

CAS: N.A.
Dataset id:7505 (Training Set)
SMILES: O=C(O)Cc1nc(N)sc1
Similarity: 0.579
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #5

CAS: N.A.
Dataset id:5869 (Training Set)
SMILES: O=C(O)C(=NOC)c1nc(N)sc1
Similarity: 0.576
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #6

CAS: N.A.
Dataset id:6698 (Training Set)
SMILES: Fc1ccc(N)c(OCC(O)C)c1(F)
Similarity: 0.575
Experimental value : Mutagenic
Predicted value : Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.637 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 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✖	Concordance for similar molecules Concordance index = 0 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.6 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

Fragment defined by the SMILES: CC(C)(F)F
The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 1 -

Prediction: Reliability:

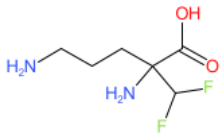
Prediction is Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)
- predicted value disagrees with experimental values of training set compounds laying in the same neuron

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogen activity: Carcinogen
P(Carcinogen): 0.621
P(NON-Carcinogen): 0.379
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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

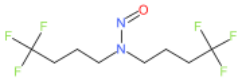
Compound #1

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.73
Experimental value : NON-Carcinogen
Predicted value : Carcinogen



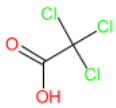
Compound #2

CAS: 70052-12-9
Dataset id:245 (Training Set)
SMILES: O=C(O)C(N)(CCCN)C(F)F
Similarity: 0.713
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



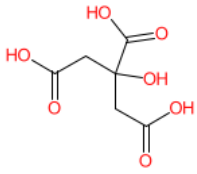
Compound #3

CAS: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.697
Experimental value : Carcinogen
Predicted value : Carcinogen



Compound #4

CAS: 76-03-9
Dataset id:761 (Test Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : NON-Carcinogen
Predicted value : Carcinogen



Compound #5

CAS: 77-92-9
Dataset id:173 (Training Set)
SMILES: O=C(O)CC(O)(C(=O)O)CC(=O)O
Similarity: 0.669
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Compound #6

CAS: 811-97-2
Dataset id:729 (Training Set)
SMILES: FCC(F)(F)F
Similarity: 0.664
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.721 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 0.491 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✖	Concordance for similar molecules Concordance index = 0 Explanation: similar molecules found in the training set have experimental values that disagree 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 = 0.6 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)..
✔	Model class assignment reliability Pos/Non-Pos difference = 0.242 Explanation: model class assignment is well defined..
⚠	Neural map neurons concordance Neurons concordance = 0.75 Explanation: predicted value disagrees with experimental values of training set compounds laying in the same neuron..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

Fragment defined by the SMILES: CC(C)(F)F
The fragment has never been found in the model's training set

1. Prediction Summary

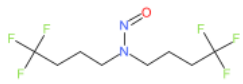
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is NON-Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogen activity: NON-Carcinogen
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

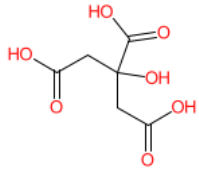
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 83335-32-4
Dataset id:496 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.697
Experimental value : Carcinogen
Predicted value : Carcinogen

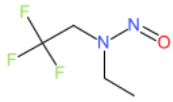
Alerts (not found also in the target): SA21 Alkyl and aryl N-nitroso groups



Compound #2

CAS: 77-92-9
Dataset id:829 (Training Set)
SMILES: O=C(O)CC(O)(C(=O)O)CC(=O)O
Similarity: 0.669
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

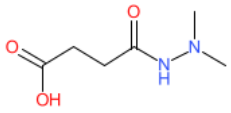
Alerts (not found also in the target): SA41 Substituted n-alkylcarboxylic acids



Compound #3

CAS: 82018-90-4
Dataset id:560 (Training Set)
SMILES: O=NN(CC)CC(F)(F)F
Similarity: 0.651
Experimental value : Carcinogen
Predicted value : Carcinogen

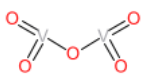
Alerts (not found also in the target): SA21 Alkyl and aryl N-nitroso groups



Compound #4

CAS: 1596-84-5
Dataset id:225 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.636
Experimental value : Carcinogen
Predicted value : Carcinogen

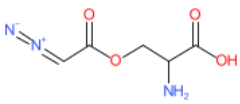
Alerts (not found also in the target): SA13 Hydrazine



Compound #5

CAS: 1314-62-1
Dataset id:771 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.632
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

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



Compound #6

CAS: 115-02-6
Dataset id:443 (Training Set)
SMILES: [N-]=[N+]=CC(=O)OCC(N)C(=O)O
Similarity: 0.629
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): SA14 Aliphatic azo and azoxy

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.35 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.682 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.517 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✖	Concordance for similar molecules Concordance index = 0.483 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.6 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

Fragment defined by the SMILES: CC(C)(F)F
The fragment has never been found in the model's training set

1. Prediction Summary

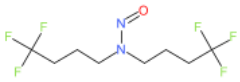
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is Possible NON-Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogenic activity: Possible NON-Carcinogen
No. alerts for carcinogenicity: 0
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

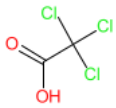
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

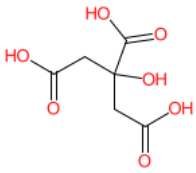
CAS: 83335-32-4
Dataset id:409 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.697
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 1; Carcinogenity alert no. 14; Carcinogenity alert no. 27



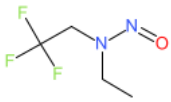
Compound #2

CAS: 76-03-9
Dataset id:912 (Training Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #3

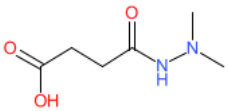
CAS: 77-92-9
Dataset id:745 (Training Set)
SMILES: O=C(O)CC(O)(C(=O)O)CC(=O)O
Similarity: 0.669
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #4

CAS: 82018-90-4
Dataset id:470 (Training Set)
SMILES: O=NN(CC)CC(F)(F)F
Similarity: 0.651
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 1; Carcinogenity alert no. 14; Carcinogenity alert no. 27

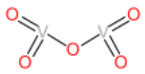


Compound #5

CAS: 1596-84-5
Dataset id:183 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.636
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 27; Carcinogenity alert no. 28

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



Compound #6

CAS: 1314-62-1
Dataset id:599 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.632
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.338 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.68 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.67 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✖	Concordance for similar molecules Concordance index = 0.324 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.6 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

Fragment defined by the SMILES: CC(C)(F)F
The fragment has never been found in the model's training set

1. Prediction Summary

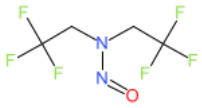
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)
The following relevant fragments have been found: Carcinogenity alert no. 125

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogenic activity: Carcinogen
No. alerts for carcinogenicity: 1
Structural Alerts: Carcinogenity alert no. 125
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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

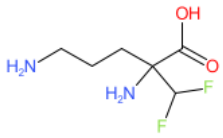


Compound #1

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.73
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 125

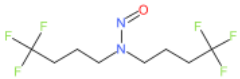
Alerts (not found also in the target): Carcinogenity alert no. 8; Carcinogenity alert no. 50; Carcinogenity alert no. 55; Carcinogenity alert no. 63



Compound #2

CAS: 70052-12-9
Dataset id:245 (Training Set)
SMILES: O=C(O)C(N)(CCCN)C(F)F
Similarity: 0.713
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 125

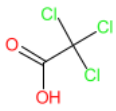


Compound #3

CAS: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.697
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 125

Alerts (not found also in the target): Carcinogenity alert no. 5; Carcinogenity alert no. 8; Carcinogenity alert no. 10; Carcinogenity alert no. 15; Carcinogenity alert no. 50; Carcinogenity alert no. 51; Carcinogenity alert no. 53; Carcinogenity alert no. 54; Carcinogenity alert no. 55; Carcinogenity alert no. 63

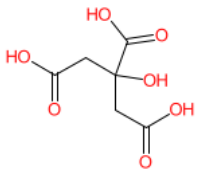


Compound #4

CAS: 76-03-9
Dataset id:761 (Test Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 57

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



Compound #5

CAS: 77-92-9
Dataset id:173 (Training Set)
SMILES: O=C(O)CC(O)(C(=O)O)CC(=O)O
Similarity: 0.669
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 21

Compound #6

CAS: 811-97-2
Dataset id:729 (Training Set)
SMILES: FCC(F)(F)F
Similarity: 0.664
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 125

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.287 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.713 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 0.322 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✖	Concordance for similar molecules Concordance index = 0.322 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.6 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on fragments/structural alerts :.

Fragment found: Carcinogenity alert no. 125

Structural alert for carcinogenity defined by the SMARTS: CCF

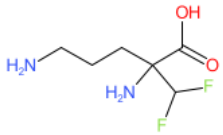
Following, the most similar compounds from the model's dataset having the same fragment.

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.73

Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 125

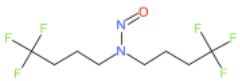
Alerts (not found also in the target): Carcinogenity alert no. 8; Carcinogenity alert no. 50; Carcinogenity alert no. 55; Carcinogenity alert no. 63



CAS: 70052-12-9
Dataset id:245 (Training Set)
SMILES: O=C(O)C(N)(CCCN)C(F)F
Similarity: 0.713

Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 125



CAS: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.697

Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 125

Alerts (not found also in the target): Carcinogenity alert no. 5; Carcinogenity alert no. 8; Carcinogenity alert no. 10; Carcinogenity alert no. 15; Carcinogenity alert no. 50; Carcinogenity alert no. 51; Carcinogenity alert no. 53; Carcinogenity alert no. 54; Carcinogenity alert no. 55; Carcinogenity alert no. 63

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

Fragment defined by the SMILES: CC(C)(F)F
The fragment has never been found in the model's training set

1. Prediction Summary

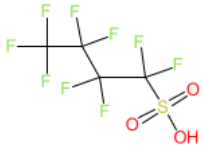
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is NON-Carcinogen, the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections.Anyway some issues could be not optimal:
- Only moderately similar compounds with known experimental value in the training set have been found

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Oral Carcinogenic class: NON-Carcinogen
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

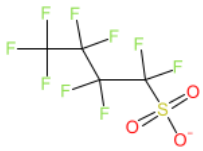


Compound #1

CAS: 375-73-5
Dataset id:627 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.785
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

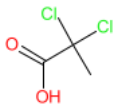
Compound #2

CAS: 335-67-1
Dataset id:629 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



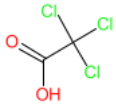
Compound #3

CAS: 29420-49-3
Dataset id:647 (Test Set)
SMILES: O=S(=O)([O-])C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #4

CAS: 75-99-0
Dataset id:424 (Test Set)
SMILES: O=C(O)C(C)(Cl)Cl
Similarity: 0.679
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 76-03-9
Dataset id:300 (Training Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : Carcinogen
Predicted value : Carcinogen

Compound #6

CAS: 811-97-2
Dataset id:690 (Training Set)
SMILES: FCC(F)(F)F
Similarity: 0.664
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✓	Global AD Index AD index = 0.883 Explanation: The predicted compound is into the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.779 Explanation: Only moderately 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:

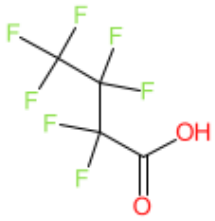
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

1. Prediction Summary

Prediction for compound Molecule 1 -

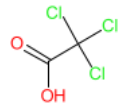


Prediction: Reliability:

Prediction is -1.37, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)

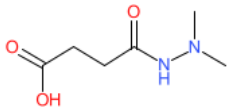
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Oral Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: -1.37
Predicted Oral Carcinogenicity SF [1/(mg/kg-day)]: 0.0426
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



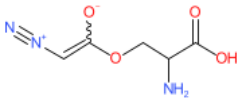
Compound #1

CAS: 76-03-9
Dataset id:300 (Training Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : -1.15
Predicted value : -1.104



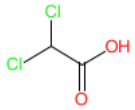
Compound #2

CAS: 1596-84-5
Dataset id:88 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.636
Experimental value : -1.74
Predicted value : -2.103



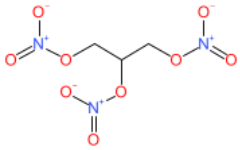
Compound #3

CAS: 115-02-6
Dataset id:27 (Training Set)
SMILES: N#[N+]C=C([O-])OCC(N)C(=O)O
Similarity: 0.629
Experimental value : 1.04
Predicted value : 0.917



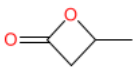
Compound #4

CAS: 79-43-6
Dataset id:109 (Training Set)
SMILES: O=C(O)C(Cl)Cl
Similarity: 0.627
Experimental value : -1.3
Predicted value : -1.01



Compound #5

CAS: 55-63-0
Dataset id:218 (Training Set)
SMILES: O=[N+]([O-])OCC(O[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.613
Experimental value : -1.77
Predicted value : -1.69



Compound #6

CAS: 3068-88-0
Dataset id:53 (Training Set)
SMILES: O=C1OC(C)C1
Similarity: 0.592
Experimental value : 0
Predicted value : -0.807

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.223 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.655 Explanation: No similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.205 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0.295 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 = 0.363 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 = 0.34 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)..

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.

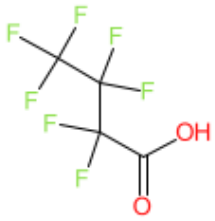
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CF The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: CC(C)(F)F The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CC(F)(F)F The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 1 -

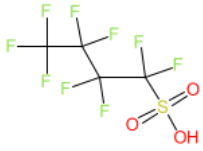


Prediction: Reliability:

Prediction is NON-Carcinogen, 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
- Accuracy of prediction for similar molecules found in the training set is not adequate
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent_fragments found)

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Inhalation Carcinogenic class: NON-Carcinogen
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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

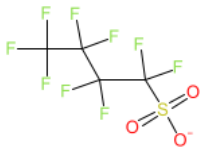


Compound #1

CAS: 375-73-5
Dataset id:613 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.785
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

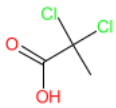
Compound #2

CAS: 335-67-1
Dataset id:615 (Test Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773
Experimental value : NON-Carcinogen
Predicted value : Carcinogen



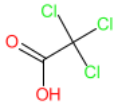
Compound #3

CAS: 29420-49-3
Dataset id:635 (Test Set)
SMILES: O=S(=O)([O-])C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #4

CAS: 75-99-0
Dataset id:387 (Training Set)
SMILES: O=C(O)C(C)(Cl)Cl
Similarity: 0.679
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 76-03-9
Dataset id:711 (Training Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

Compound #6

CAS: 811-97-2
Dataset id:685 (Test Set)
SMILES: FCC(F)(F)F
Similarity: 0.664
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.632 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.779 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 0.505 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✔	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 = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent_fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

Fragment defined by the SMILES: CC(C)(F)F
The fragment has less than 3 occurrences in the model's training set

1. Prediction Summary

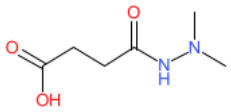
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is -0.32, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (3 unknown fragments found)

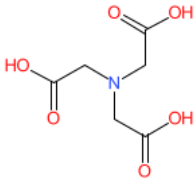
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Inhalation Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: -0.32
Predicted Inhalation Carcinogenicity SF [1/(mg/kg-day)]: 0.476
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



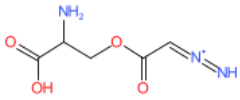
Compound #1

CAS: 1596-84-5
Dataset id:74 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.636
Experimental value : -1.75
Predicted value : -0.976



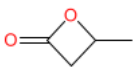
Compound #2

CAS: 139-13-9
Dataset id:178 (Training Set)
SMILES: O=C(O)CN(CC(=O)O)CC(=O)O
Similarity: 0.624
Experimental value : -2.28
Predicted value : 0.003



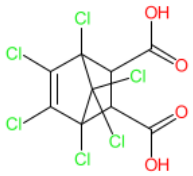
Compound #3

CAS: 115-02-6
Dataset id:24 (Training Set)
SMILES: O=C(O)C(N)COC(=O)C=[N+]=N
Similarity: 0.622
Experimental value : 1.04
Predicted value : -0.661



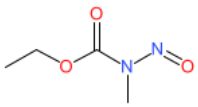
Compound #4

CAS: 3068-88-0
Dataset id:45 (Training Set)
SMILES: O=C1OC(C)C1
Similarity: 0.592
Experimental value : 0.01
Predicted value : 0.684



Compound #5

CAS: 115-28-6
Dataset id:53 (Test Set)
SMILES: O=C(O)C1C(C(=O)O)C2(C(=C(C1(C2(Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.589
Experimental value : -1.04
Predicted value : 1.623



Compound #6

CAS: 615-53-2
Dataset id:191 (Training Set)
SMILES: O=NN(C(=O)OCC)C
Similarity: 0.583
Experimental value : 2.04
Predicted value : 0.859

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.252 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.63 Explanation: No similar compounds with known experimental value in the training set have been found..
✗	Accuracy of prediction for similar molecules Accuracy index = 1.529 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✗	Concordance for similar molecules Concordance index = 1.693 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✗	Maximum error of prediction among similar molecules Max error index = 2.283 Explanation: the maximum error in prediction of similar molecules found in the training set has a high 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 = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (3 unknown fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CC(C)(F)F The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CC(F)(F)F The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CF The fragment has never been found in the model's training set

1. Prediction Summary

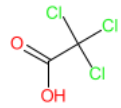
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is 0.3091, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)

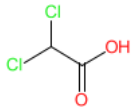
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: 0.3091
Predicted TD50 [mg/kg bw/day]: 0.4908
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



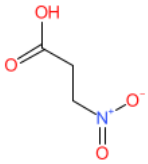
Compound #1

CAS: N.A.
Dataset id:45 (Training Set)
SMILES: C(C(=O)O)(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : -3.827
Predicted value : -3.468



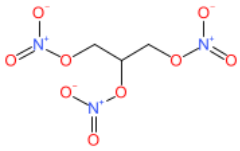
Compound #2

CAS: N.A.
Dataset id:25 (Training Set)
SMILES: C(C(=O)O)(Cl)Cl
Similarity: 0.627
Experimental value : -2.422
Predicted value : -3.345



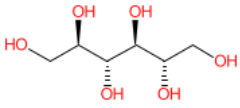
Compound #3

CAS: N.A.
Dataset id:6 (Training Set)
SMILES: N(=O)(=O)CCC(=O)O
Similarity: 0.617
Experimental value : -0.644
Predicted value : -1.254



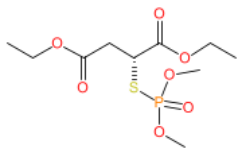
Compound #4

CAS: N.A.
Dataset id:24 (Training Set)
SMILES: C(C(CON(=O)=O)ON(=O)=O)ON(=O)=O
Similarity: 0.613
Experimental value : -2.344
Predicted value : -0.701



Compound #5

CAS: N.A.
Dataset id:162 (Test Set)
SMILES: C([C@@H]([C@@H]([C@@H]([C@@H](CO)O)O)O)O)O
Similarity: 0.598
Experimental value : -4.255
Predicted value : -3.919



Compound #6

CAS: N.A.
Dataset id:38 (Training Set)
SMILES: CCOC(=O)[C@@H](CC(=O)OCC)SP(=O)(OC)OC
Similarity: 0.595
Experimental value : -3.193
Predicted value : -3.699

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.221 Explanation: The predicted compound is outside the Applicability Domain of the model.
✖	Similar molecules with known experimental value Similarity index = 0.65 Explanation: No similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.641 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 3.434 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Maximum error of prediction among similar molecules Max error index = 0.923 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate 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 = 0.34 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CF The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: CC(C)(F)F The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CC(F)(F)F The fragment has never been found in the model's training set

1. Prediction Summary

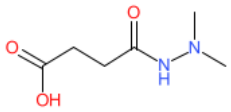
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is -3.7781, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)

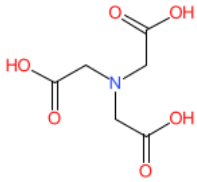
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: -3.7781
Predicted TD50 [mg/kg bw/day]: 5999.35
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



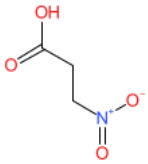
Compound #1

CAS: N.A.
Dataset id:111 (Training Set)
SMILES: C(=O)(O)CCC(=O)NN(C)C
Similarity: 0.636
Experimental value : -3.984
Predicted value : -2.932



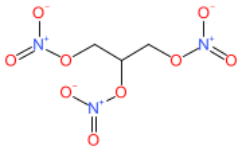
Compound #2

CAS: N.A.
Dataset id:98 (Training Set)
SMILES: N(CC(=O)O)(CC(=O)O)CC(=O)O
Similarity: 0.624
Experimental value : -2.908
Predicted value : -2.721



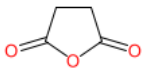
Compound #3

CAS: N.A.
Dataset id:156 (Test Set)
SMILES: N(=O)(=O)CCC(=O)O
Similarity: 0.617
Experimental value : -1.462
Predicted value : -2.607



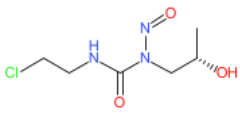
Compound #4

CAS: N.A.
Dataset id:94 (Training Set)
SMILES: C(C(CON(=O)=O)ON(=O)=O)ON(=O)=O
Similarity: 0.613
Experimental value : -2.517
Predicted value : -6.738



Compound #5

CAS: N.A.
Dataset id:45 (Training Set)
SMILES: C1CC(=O)OC1=O
Similarity: 0.596
Experimental value : -3.114
Predicted value : -4.71



Compound #6

CAS: N.A.
Dataset id:152 (Test Set)
SMILES: N(C[C@@H](O)C)(C(=O)NCCCl)N=O
Similarity: 0.592
Experimental value : 0.053
Predicted value : -0.968

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.214 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.63 Explanation: No similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.619 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0.538 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 = 1.052 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate 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 = 0.34 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)..

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.

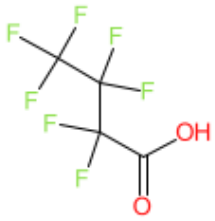
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CF The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: CC(C)(F)F The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CC(F)(F)F The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 1 -



EXPERIMENTAL DATA

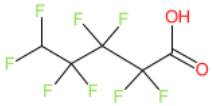
E xperimental value is 1414.33 mg/Kg. Model prediction is 1414.33 mg/kg (GOOD reliability).

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: 0.82
Predicted log LD50 [log(mmol/Kg)]: 0.82
Predicted log LD50 [mg/Kg]: 1414.33
Molecules used for prediction: 1
Experimental value [mg/Kg]: 1414.33
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

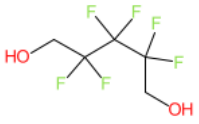
Compound #1

CAS: N.A.
Dataset id:3430 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 1
Experimental value : 0.82
Predicted value : 0.808



Compound #2

CAS: N.A.
Dataset id:3530 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)F
Similarity: 0.927
Experimental value : 0.85
Predicted value : 0.786



Compound #3

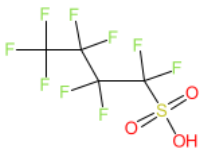
CAS: N.A.
Dataset id:3547 (Training Set)
SMILES: FC(F)(CO)C(F)(F)C(F)(F)CO
Similarity: 0.856
Experimental value : 0.92
Predicted value : 0.835

Compound #4

CAS: N.A.
Dataset id:4123 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)F
Similarity: 0.82
Experimental value : 0.62
Predicted value : 0.397

Compound #5

CAS: N.A.
Dataset id:3089 (Training Set)
SMILES: O=C(C(F)(F)F)C(F)(F)F
Similarity: 0.803
Experimental value : -0.06
Predicted value : 0.082



Compound #6

CAS: N.A.
Dataset id:3431 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.785
Experimental value : 0.16
Predicted value : -

3.2 Applicability Domain:
Measured Applicability Domain Scores

✓	Global AD Index AD index = 1 Explanation: The predicted compound is into the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 1 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.012 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0 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 = 0.012 Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability..
✓	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.

1. Prediction Summary

Prediction for compound Molecule 1 -

Prediction: Reliability:

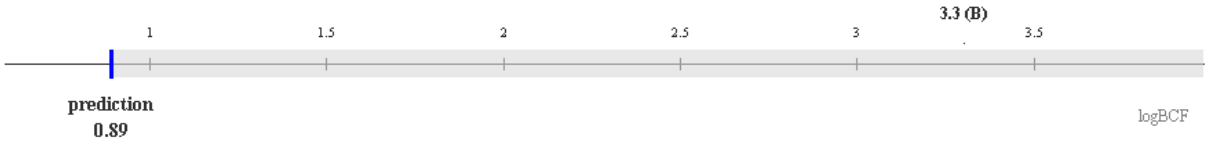
Prediction is 0.89 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability
The following relevant fragments have been found: Carbonyl residue (SR 02); COOH group (PG 01)

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF [log(L/kg)]: 0.89
Predicted BCF [L/kg]: 8
Predicted BCF from sub-model 1 (HM) [log(L/kg)]: 1.15
Predicted BCF from sub-model 2 (GA) [log(L/kg)]: 1.02
Predicted LogP (MLogP): 1.9
Structural Alerts: Carbonyl residue (SR 02); COOH group (PG 01)
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

2. Possible Use and Uncertainty

Threshold 3.3 (bioaccumulative)

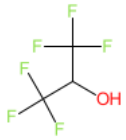
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.3, the current compound can not be associated (due to its Applicability Domain index value) to any conservative interval.
No safe classification can be done.



Threshold 3.7 (very bioaccumulative)

Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.7, the current compound can not be associated (due to its Applicability Domain index value) to any conservative interval.
No safe classification can be done.

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



Compound #1

CAS: 920-66-1
Dataset id:263 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833
Experimental value : 0.3
Predicted value : 0.601

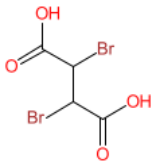
Alerts (not found also in the target): OH group (PG 06)

Compound #2

CAS: 335-67-1
Dataset id:56 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773
Experimental value : 3.12
Predicted value : 2.534

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

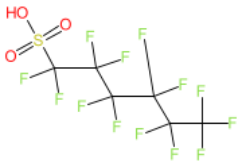
Alerts (not found also in the target): 10 F atoms in the molecule (SO 10)



Compound #3

CAS: 526-78-3
Dataset id:71 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.704
Experimental value : 0.81
Predicted value : 0.227

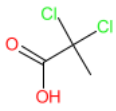
Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)



Compound #4

CAS: 355-46-4
Dataset id:55 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.703
Experimental value : 3.6
Predicted value : 1.585

Alerts (not found also in the target): 10 F atoms in the molecule (SO 10); SO3H group (PG 02)

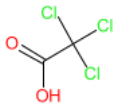


Compound #5

CAS: 75-99-0
Dataset id:468 (Test Set)
SMILES: O=C(O)C(C)(Cl)Cl
Similarity: 0.679
Experimental value : 0.85
Predicted value : 0.075

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

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



Compound #6

CAS: 76-03-9
Dataset id:397 (Test Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : -0.15
Predicted value : -0.038

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.75 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.8 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.443 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 1.41 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Maximum error of prediction among similar molecules Max error index = 0.586 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate 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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on fragments/structural alerts - 1 of 2:.

Fragment found: Carbonyl residue (SR 02)

This chemical contains a carbonyl residue. This residue has been found to be present in a very large (112) number of non-bioaccumulative compounds, even when the logP value was higher than 3.

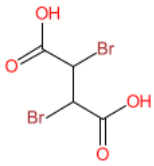
Following, the most similar compounds from the model's dataset having the same fragment.

CAS: 335-67-1
Dataset id:56 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773

Experimental value : 3.12
Predicted value : 2.534

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

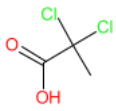
Alerts (not found also in the target): 10 F atoms in the molecule (SO 10)



CAS: 526-78-3
Dataset id:71 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.704

Experimental value : 0.81
Predicted value : 0.227

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)



CAS: 75-99-0
Dataset id:468 (Test Set)
SMILES: O=C(O)C(C)(Cl)Cl
Similarity: 0.679

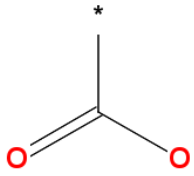
Experimental value : 0.85
Predicted value : 0.075

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 1) Reasoning on fragments/structural alerts - 2 of 2:.

Fragment found: COOH group (PG 01)



This chemical contains a COOH polar group.The presence of polar groups increases hydrophilicity, related to lower values of BCF.

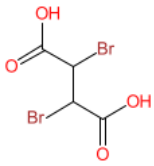
Following, the most similar compounds from the model's dataset having the same fragment.

CAS: 335-67-1
Dataset id:56 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773

Experimental value : 3.12
Predicted value : 2.534

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

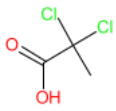
Alerts (not found also in the target): 10 F atoms in the molecule (SO 10)



CAS: 526-78-3
Dataset id:71 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.704

Experimental value : 0.81
Predicted value : 0.227

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)



CAS: 75-99-0
Dataset id:468 (Test Set)
SMILES: O=C(O)C(C)(Cl)Cl
Similarity: 0.679

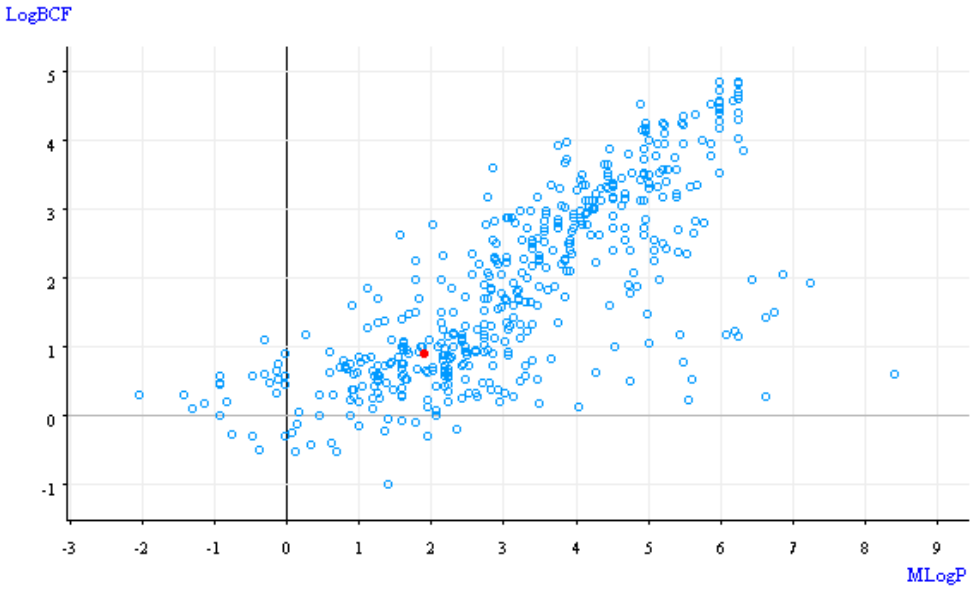
Experimental value : 0.85
Predicted value : 0.075

Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

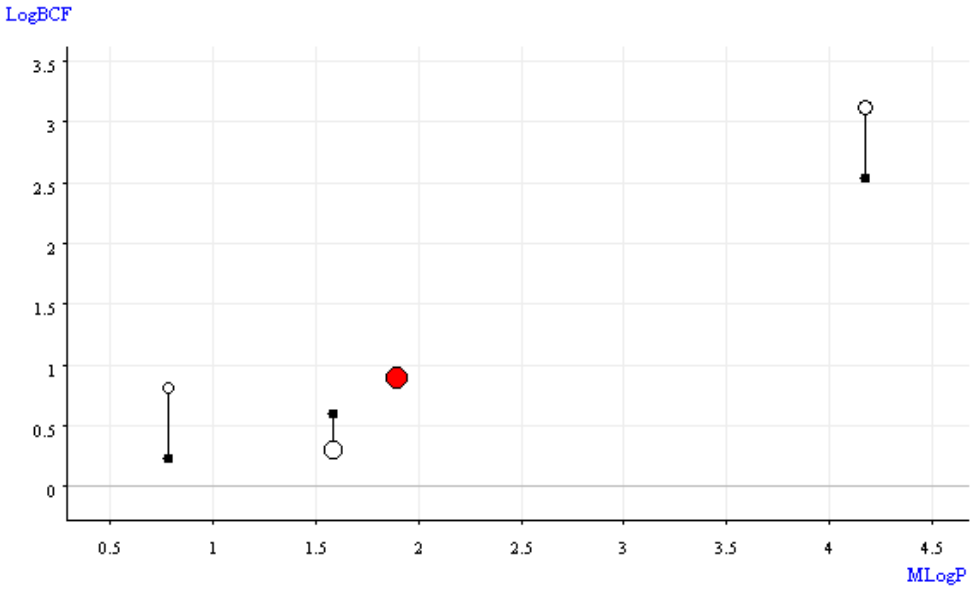
4.2 Reasoning:
Analysis of Molecular Descriptors

Descriptor name: MLogP
Descriptor name: LogP is directly correlated to the logBCF value.

Following, a scatterplot of MLogP against response values; experimental values are reported for the training set, predicted value for the studied compound. Light blue dots represent values of compounds from training set, red dot is the value of the studied compound.

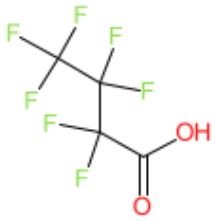


Following, a scatterplot of MLogP against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound; the size of the circle is proportional to the similarity to the studied compound.



1. Prediction Summary

Prediction for compound Molecule 1 -

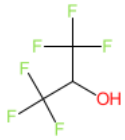


Prediction: Reliability:

Prediction is 0.5 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)

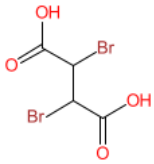
Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF [log(L/kg)]: 0.5
Predicted BCF [L/kg]: 3
Predicted LogP (Meylan/Kowwin): 2.43
Predicted LogP reliability: Good
MW: 213.8
Ionic compound: yes
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



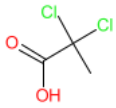
Compound #1

CAS: 920-66-1
Dataset id:117 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833
Experimental value : 0.4
Predicted value : 0.762



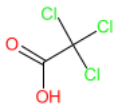
Compound #2

CAS: 526-78-3
Dataset id:27 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.704
Experimental value : 0.81
Predicted value : 0.5



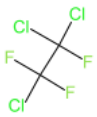
Compound #3

CAS: 127-20-8
Dataset id:20 (Training Set)
SMILES: O=C(O)C(C)(Cl)Cl
Similarity: 0.679
Experimental value : 0.86
Predicted value : 0.5



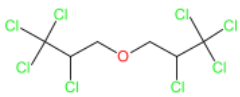
Compound #4

CAS: 76-03-9
Dataset id:5 (Training Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : 0.23
Predicted value : 0.5



Compound #5

CAS: 76-13-1
Dataset id:668 (Test Set)
SMILES: FC(F)(C(F)(Cl)Cl)Cl
Similarity: 0.646
Experimental value : 1.7
Predicted value : 1.752



Compound #6

CAS: 127-90-2
Dataset id:406 (Training Set)
SMILES: O(CC(C(Cl)(Cl)Cl)Cl)CC(C(Cl)(Cl)Cl)Cl
Similarity: 0.631
Experimental value : 3.28
Predicted value : 3.034

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.385 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.756 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.336 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0.205 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 = 0.362 Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability..
✓	Reliability of logP prediction LogP reliability = 1 Explanation: reliability of logP value used by the model is good..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the defined range..
✗	Atom Centered Fragments similarity check ACF index = 0.51 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)..

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.

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

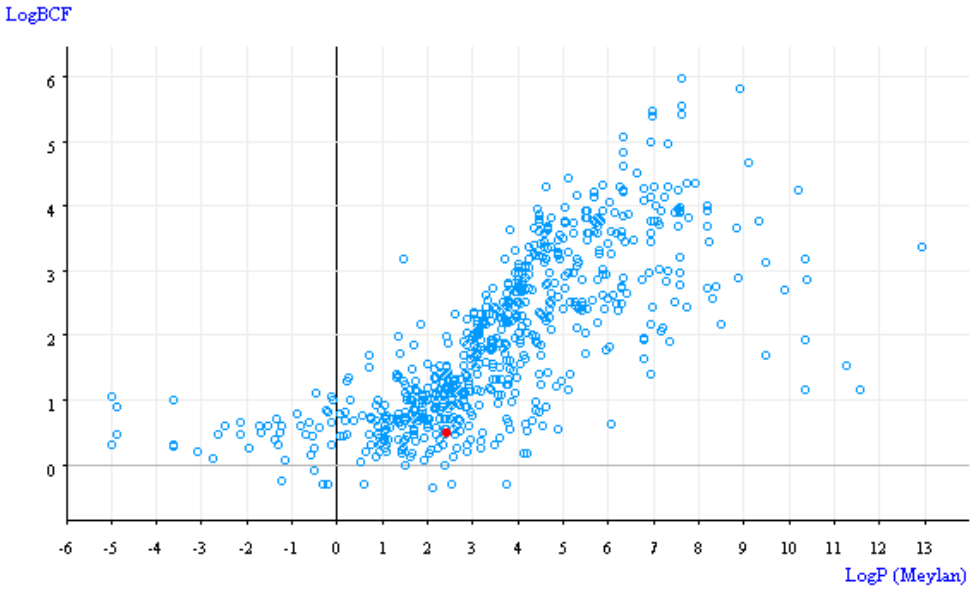
(Molecule 1) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CC(F)(F)F The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: CC(C)(F)F The fragment has never been found in the model's training set

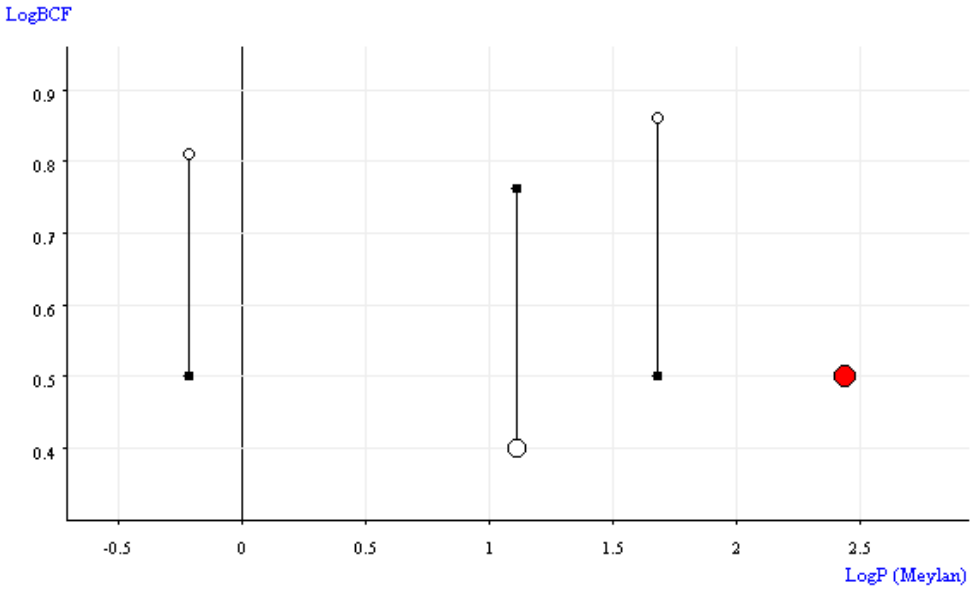
4.2 Reasoning:
Analysis of Molecular Descriptors

Descriptor name: LogP (Meylan)
Descriptor name: LogP is directly correlated to the logBCF value.

Following, a scatterplot of LogP (Meylan) against response values; experimental values are reported for the training set, predicted value for the studied compound. Light blue dots represent values of compounds from training set, red dot is the value of the studied compound.



Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound; the size of the circle is proportional to the similarity to the studied compound.



1. Prediction Summary

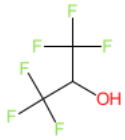
Prediction for compound Molecule 1 -

Prediction: Reliability:

Prediction is 1.45 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF (up) [log(L/kg)]: 1.45
Predicted BCF (up) [L/kg]: 28
Predicted BCF (low) [log(L/kg)]: 1.23
Predicted BCF (low) [L/kg]: 17
Predicted BCF (mid) [log(L/kg)]: 1.28
Predicted BCF (mid) [L/kg]: 19
Predicted LogP (Meylan/Kowwin): 2.43
Predicted LogP reliability: Good
Predicted kM (Meylan): 0.13
Predicted kM reliability: Low
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



Compound #1

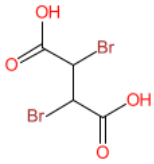
CAS: 920-66-1
Dataset id:191 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833
Experimental value : 0.2
Predicted value : 0.565

Compound #2

CAS: 335-67-1
Dataset id:644 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773
Experimental value : 0.977
Predicted value : 4

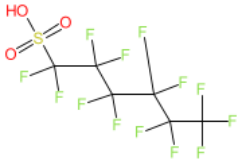
Compound #3

CAS: 335-76-2
Dataset id:288 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.715
Experimental value : 3.04
Predicted value : 3.236



Compound #4

CAS: 526-78-3
Dataset id:477 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.704
Experimental value : 0.145
Predicted value : -0.032



Compound #5

CAS: 3871-99-6
Dataset id:430 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.703
Experimental value : 1.62
Predicted value : 2.989

Compound #6

CAS: 2058-94-8
Dataset id:665 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.692
Experimental value : 3.72
Predicted value : 2.565

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.75 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.8 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 1.694 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
⚠	Concordance for similar molecules Concordance index = 0.86 Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Maximum error of prediction among similar molecules Max error index = 3.023 Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability..
✔	Reliability of logP prediction LogP reliability = 1 Explanation: reliability of logP value used by the model is good..
✔	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.

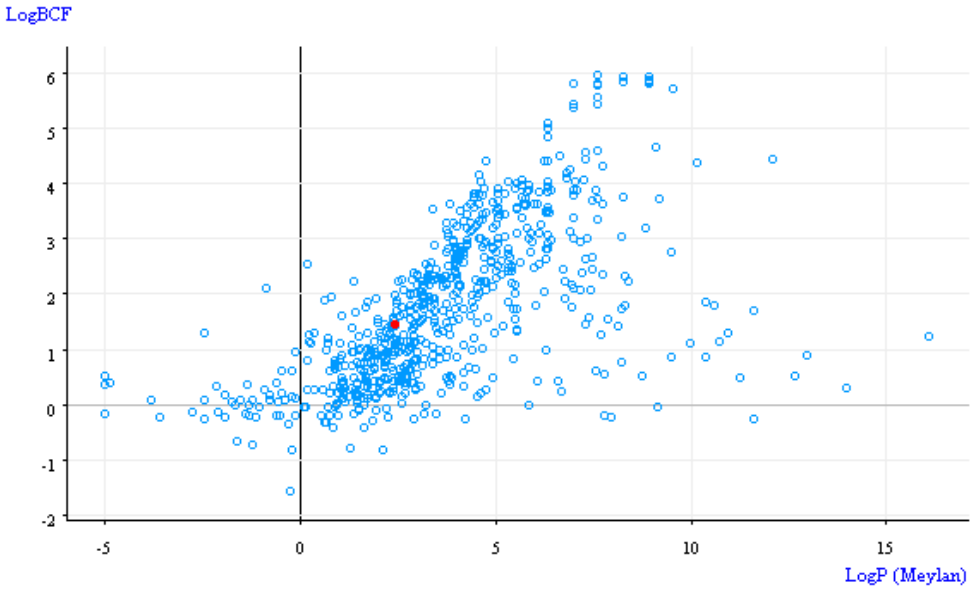
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

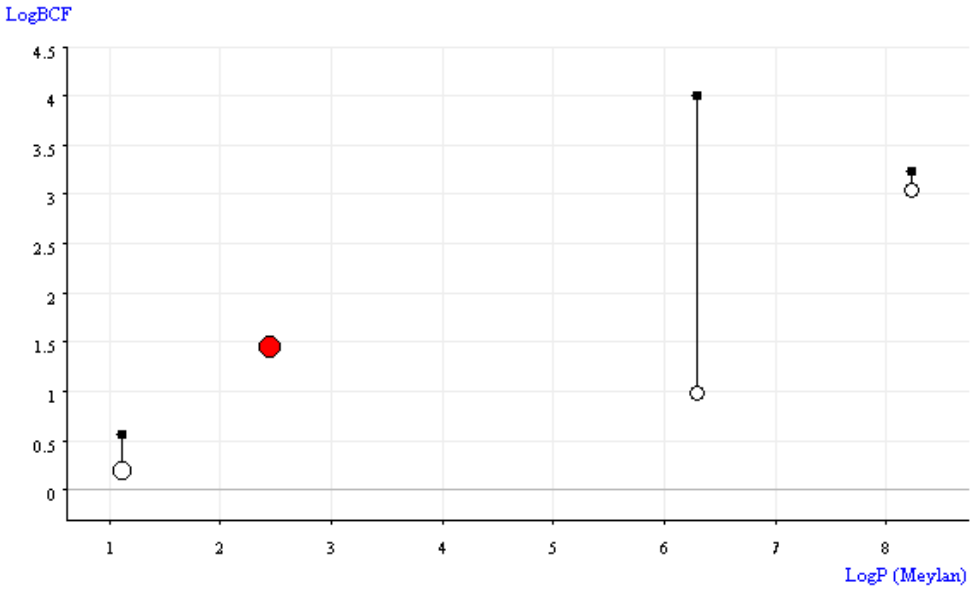
4.2 Reasoning:
Analysis of Molecular Descriptors

Descriptor name: LogP (Meylan)
Descriptor name: LogP is directly correlated to the logBCF value.

Following, a scatterplot of LogP (Meylan) against response values; experimental values are reported for the training set, predicted value for the studied compound. Light blue dots represent values of compounds from training set, red dot is the value of the studied compound.

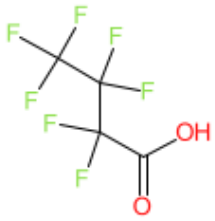


Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound; the size of the circle is proportional to the similarity to the studied compound.



1. Prediction Summary

Prediction for compound Molecule 1 -

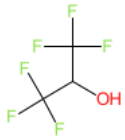


Prediction: Reliability:

Prediction is 1.72 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability

Compound: Molecule 1
Compound SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF [log(L/kg)]: 1.72
Molecules used for prediction: 4
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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

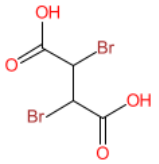


Compound #1

CAS: 920-66-1
Dataset id:391 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.833
Experimental value : 0.244
Predicted value : 1.753

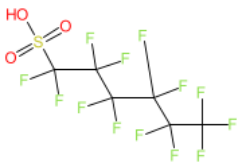
Compound #2

CAS: 335-67-1
Dataset id:308 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.773
Experimental value : 3.12
Predicted value : 3.096



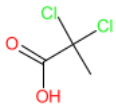
Compound #3

CAS: 526-78-3
Dataset id:325 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.704
Experimental value : 0.43
Predicted value : 0.02



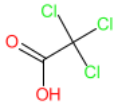
Compound #4

CAS: 355-46-4
Dataset id:310 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.703
Experimental value : 3.6
Predicted value : 3.479



Compound #5

CAS: 75-99-0
Dataset id:40 (Training Set)
SMILES: O=C(O)C(C)(Cl)Cl
Similarity: 0.679
Experimental value : 0.664
Predicted value : 0.188



Compound #6

CAS: 76-03-9
Dataset id:41 (Training Set)
SMILES: O=C(O)C(Cl)(Cl)Cl
Similarity: 0.677
Experimental value : 0.025
Predicted value : 0.409

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.7 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.741 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.516 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 1.511 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Maximum error of prediction among similar molecules Max error index = 1.509 Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability..
✔	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.

References and Documentation

You can find complete details on each model and on how to read results in the proper model's guide, available on-line at www.vega-qsar.eu or directly in the VegaNIC application.

Mutagenicity (Ames test) CONSENSUS model(version 1.0.4)

Mutagenicity (Ames test) Consensus model based on the predictions of VEGA mutagenicity models.

Mutagenicity (Ames test) model (CAESAR)(version 2.1.14)

QSAR classification model for Mutagenicity (from CAESAR project)

Mutagenicity (Ames test) model (ISS)(version 1.0.3)

Classification model for Mutagenicity (Ames test) based on Benigni-Bossa (Istituto Superiore di Sanità) rule set

Mutagenicity (Ames test) model (SarPy-IRFMN)(version 1.0.8)

QSAR classification model for Mutagenicity (SarPy/IRFMN)

References and Documentation

Mutagenicity (Ames test) model (KNN-Read-Across)(version 1.0.1)

KNN (Read-Across) model for Mutagenicity (Ames test)

Mutagenicity (Ames test) model for aromatic amines (CONCERT/IRFMN)(version 1.0.0)

A knowledge-based expert rule system for predicting mutagenicity (AMES test) of aromatic amines.

Carcinogenicity model (CAESAR)(version 2.1.10)

QSAR classification model for Carcinogenicity (from CAESAR project)

Carcinogenicity model (ISS)(version 1.0.3)

Classification model for Carcinogenicity based on Benigni-Bossa (Istituto Superiore di Sanità) rule set

References and Documentation

Carcinogenicity model (IRFMN-ISSCAN-CGX)(version 1.0.2)

QSAR classification model for Carcinogenicity (IRFMN/ISSCAN-CGX) based on the ISSCAN-CGX dataset

Carcinogenicity model (IRFMN-Antares)(version 1.0.2)

QSAR classification model for Carcinogenicity (IRFMN/Antares) based on the Antares dataset

Carcinogenicity oral classification model (IRFMN)(version 1.0.1)

Classification model for carcinogenicity (oral route).

Carcinogenicity oral Slope Factor model (IRFMN)(version 1.0.1)

Quantitative model for the carcinogenicity (oral route) Slope Factor.

References and Documentation

Carcinogenicity inhalation classification model (IRFMN)(version 1.0.1)

Classification model for carcinogenicity (inhalation route).

Carcinogenicity inhalation Slope Factor model (IRFMN)(version 1.0.1)

Quantitative model for the carcinogenicity inhalation route) Slope Factor.

Carcinogenicity in male rat (CORAL)(version 1.0.0)

Carcinogenicity in male rat quantitative (TD50) model model (CORAL)

Carcinogenicity in female Rat (CORAL)(version 1.0.0)

Carcinogenicity in female rat quantitative (TD50) model model (CORAL)

References and Documentation

Acute Toxicity (LD50) model (KNN)(version 1.0.0)

KNN model for acute toxicity (LD50)

BCF model (CAESAR)(version 2.1.15)

QSAR regression model for fish BCF (from CAESAR project)

BCF model (Meylan)(version 1.0.4)

QSAR regression model for fish BCF (based on Meylan model)

BCF model (Arnot-Gobas)(version 1.0.1)

QSAR regression model for fish BCF (based on Arnot-Gobas model)

References and Documentation

BCF model (KNN-Read-Across)(version 1.1.1)

KNN (Read-Across) model for fish BCF