VEGA QSAR MODELS

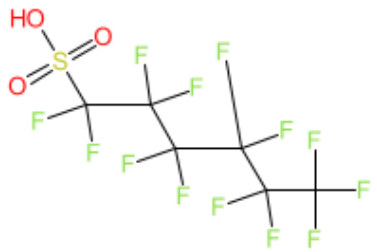
Report

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 0 -



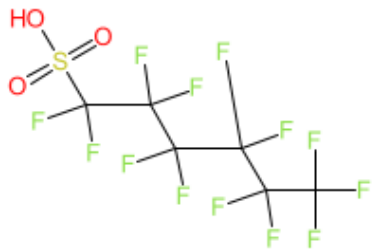
Prediction:

Prediction is NON-Mutagenic with a consensus score of 0.3, based on 4 models.

Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)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: 0.3
Model Caesar assessment: NON-Mutagenic (LOW reliability)
Model ISS assessment: NON-Mutagenic (LOW reliability)
Model SarPy assessment: NON-Mutagenic (LOW reliability)
Model KNN assessment: NON-Mutagenic (MODERATE reliability)
Remarks:
none

1. Prediction Summary

Prediction for compound Molecule 0 -

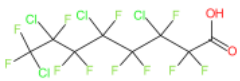


Prediction: Reliability:

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
- 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)

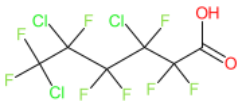
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



Compound #1

CAS: 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: 0.787
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

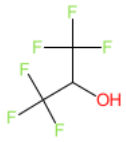


Compound #2

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: 0.774
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

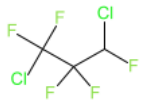
Compound #3

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: 0.769
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

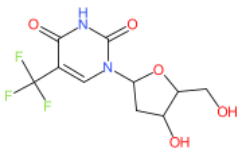
CAS: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.638
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #5

CAS: 507-55-1
Dataset id:34 (Training Set)
SMILES: FC(C(F)(F)C(F)(F)Cl)Cl
Similarity: 0.607
Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

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



Compound #6

CAS: 70-00-8
Dataset id:4136 (Training Set)
SMILES: O=C1NC(=O)N(C=C1C(F)(F)F)C2OC(CO)C(O)C2
Similarity: 0.605
Experimental value : Mutagenic
Predicted value : Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.529 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.776 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 = 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.

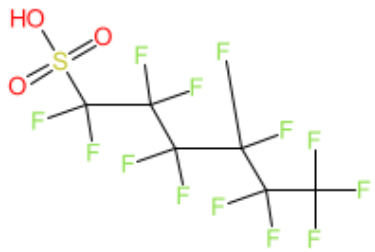
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S
The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

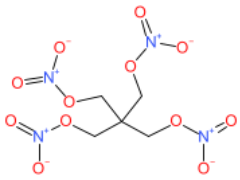


Prediction: Reliability:

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:
- No similar compounds with known experimental value in the training set have been found
- some 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 (2 unknown fragments and 1 infrequent_fragments found)

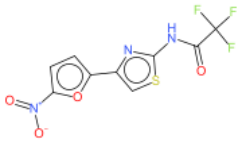
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



Compound #1

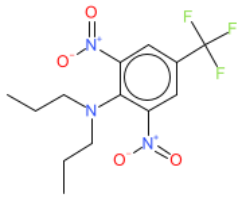
CAS: 78-11-5
Dataset id:617 (Training Set)
SMILES: O=[N+]([O-])OCC(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.591
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #2

CAS: 42011-48-3
Dataset id:763 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.587
Experimental value : Mutagenic
Predicted value : Mutagenic

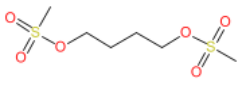
Alerts (not found also in the target): SA27 Nitro aromatic



Compound #3

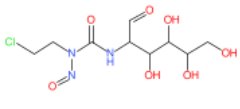
CAS: 1582-09-8
Dataset id:234 (Training Set)
SMILES: O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F
Similarity: 0.579
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA27 Nitro aromatic



Compound #4

CAS: 55-98-1
Dataset id:249 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.577
Experimental value : Mutagenic
Predicted value : NON-Mutagenic

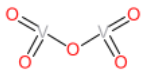


Compound #5

CAS: 54749-90-5
Dataset id:826 (Training Set)
SMILES: O=NN(C(=O)NC(C=O)C(O)C(O)C(O)CO)CCCl
Similarity: 0.576
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA8 Aliphatic halogens; SA11 Simple aldehyde; SA21 Alkyl and aryl N-nitroso groups

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



Compound #6

CAS: 1314-62-1
Dataset id:771 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.566
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.22 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.589 Explanation: No similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
⚠	Concordance for similar molecules Concordance index = 0.504 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 = 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.

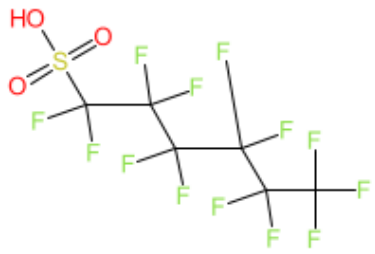
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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
	Fragment defined by the SMILES: CC(F)(F)S The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

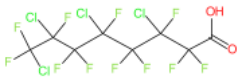


Prediction: Reliability:

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 adequate
- 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: SM150; SM153

Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)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: 2
Structural Alerts: SM150; SM153
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

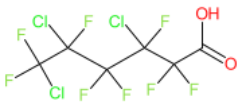


Compound #1

CAS: 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: 0.787
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

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



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: 0.774
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

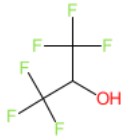
Alerts (not found also in the target): SM106; SM177

Compound #3

CAS: 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: 0.769
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM150

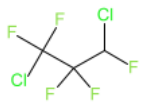
Alerts (not found also in the target): SM143; SM163; SM177



Compound #4

CAS: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.638
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM150



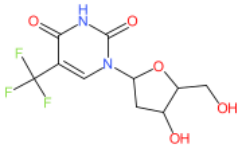
Compound #5

CAS: 507-55-1
Dataset id:34 (Training Set)
SMILES: FC(C(F)(F)C(F)(F)Cl)Cl
Similarity: 0.607
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM106

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



Compound #6

CAS: 70-00-8
Dataset id:4136 (Training Set)
SMILES: O=C1NC(=O)N(C=C1C(F)(F)F)C2OC(CO)C(O)C2
Similarity: 0.605
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM94; SM98; SM156

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.4 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.776 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✗	Accuracy of prediction for similar molecules Accuracy index = 0.329 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 = 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.

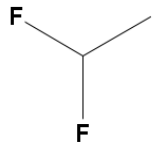
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

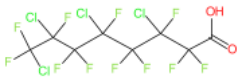
(Molecule 0) Reasoning on fragments/structural alerts - 1 of 2:.

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.

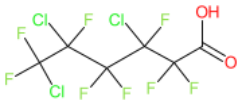


CAS: 2923-68-4
Dataset id:2832 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl)Cl
Similarity: 0.787

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

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



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: 0.774

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

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

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: 0.769

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

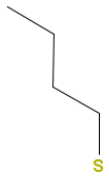
Alerts (found also in the target): SM150

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

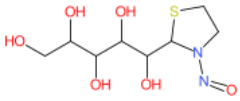
(Molecule 0) Reasoning on fragments/structural alerts - 2 of 2:.

Fragment found: SM153



Sarpy alert n. 153 for NON-Mutagenicity, defined by SMARTS: SCCCC

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

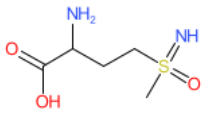


CAS: 92134-96-8
Dataset id:1424 (Training Set)
SMILES: O=NN1CCSC1C(O)C(O)C(O)C(O)CO
Similarity: 0.578

Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM153

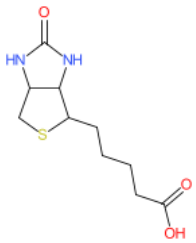
Alerts (not found also in the target): SM2; SM103



CAS: 1982-67-8
Dataset id:309 (Training Set)
SMILES: O=C(O)C(N)CCS(=O)(=N)C
Similarity: 0.556

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

Alerts (found also in the target): SM153



CAS: 58-85-5
Dataset id:4120 (Training Set)
SMILES: O=C1NC2CSC(CCCCC(=O)O)C2(N1)
Similarity: 0.551

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

Alerts (found also in the target): SM153

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

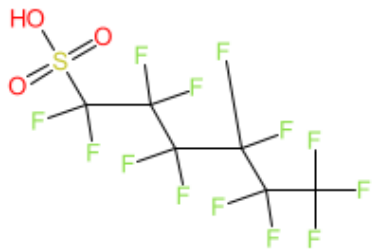
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S
The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

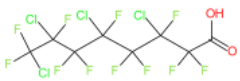


Prediction: Reliability:

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

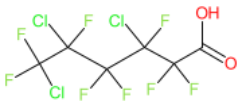
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Molecules used for prediction: 4
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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



Compound #1

CAS: 2923-68-4
Dataset id:2605 (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: 0.787
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

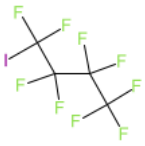


Compound #2

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: 0.774
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

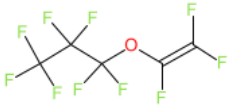
Compound #3

CAS: 335-76-2
Dataset id:2785 (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: 0.769
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: 0.702
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



Compound #5

CAS: 1623-05-8
Dataset id:5707 (Training Set)
SMILES: FC(F)=C(F)OC(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.693
Experimental value : Mutagenic
Predicted value : NON-Mutagenic

Compound #6

CAS: 354-87-0
Dataset id:5751 (Training Set)
SMILES: O=S(=O)(F)C(F)(F)C(F)(F)F
Similarity: 0.69
Experimental value : Mutagenic
Predicted value : Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.814 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.753 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.776 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 = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

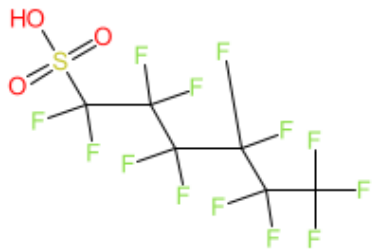
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 0 -

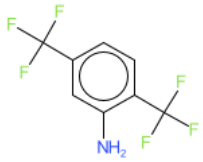


Prediction: Reliability:

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:
- 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
- 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)

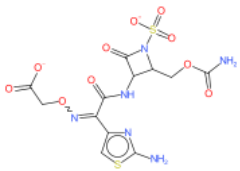
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NA
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



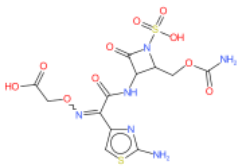
Compound #1

CAS: N.A.
Dataset id:7541 (Training Set)
SMILES: FC(F)(F)c1ccc(c(N)c1)C(F)(F)F
Similarity: 0.58
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



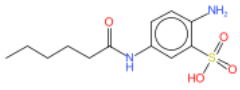
Compound #2

CAS: N.A.
Dataset id:5148 (Training Set)
SMILES: O=C(OCC2N(C(=O)C2(NC(=O)C(=NOCC(=O)[O-])c1nc(N)sc1))S(=O)(=O)[O-])N
Similarity: 0.571
Experimental value : NON-Mutagenic
Predicted value : NA



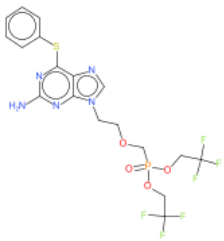
Compound #3

CAS: N.A.
Dataset id:5147 (Training Set)
SMILES: O=C(OCC2N(C(=O)C2(NC(=O)C(=NOCC(=O)O)c1nc(N)sc1))S(=O)(=O)O)N
Similarity: 0.567
Experimental value : NON-Mutagenic
Predicted value : NA



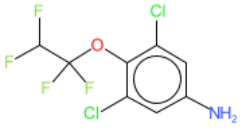
Compound #4

CAS: N.A.
Dataset id:5450 (Training Set)
SMILES: O=C(Nc1ccc(N)c(c1)S(=O)(=O)O)CCCCC
Similarity: 0.566
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



Compound #5

CAS: N.A.
Dataset id:7303 (Training Set)
SMILES: O=P(OCC(F)(F)F)(OCC(F)(F)F)COCCn3cnc1c3(nc(nc1Sc2ccccc2)N)
Similarity: 0.562
Experimental value : NON-Mutagenic
Predicted value : NA



Compound #6

CAS: N.A.
Dataset id:5582 (Training Set)
SMILES: FC(F)C(F)(F)Oc1c(cc(N)cc1Cl)Cl
Similarity: 0.56
Experimental value : NON-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 = 0.575 Explanation: No 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 = 0.4 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:

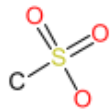
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

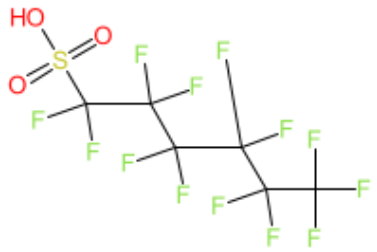
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CS(O)(=O)=O The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

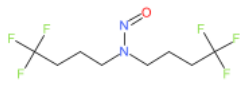


Prediction: Reliability:

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 optimal
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- 1descriptor(s) for this compound have values outside the descriptor range of the compounds of the training set.
- 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)
- predicted substance falls into a neuron that is populated by no compounds of the training set

Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogen activity: Carcinogen
P(Carcinogen): 0.736
P(NON-Carcinogen): 0.264
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

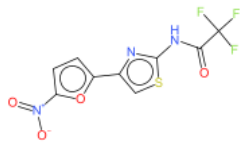


Compound #1

CAS: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.654
Experimental value : Carcinogen
Predicted value : Carcinogen

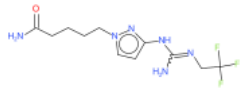
Compound #2

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.612
Experimental value : NON-Carcinogen
Predicted value : Carcinogen



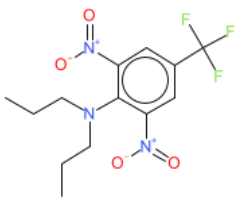
Compound #3

CAS: 42011-48-3
Dataset id:774 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.587
Experimental value : Carcinogen
Predicted value : Carcinogen



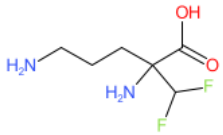
Compound #4

CAS: 84545-30-2
Dataset id:392 (Training Set)
SMILES: O=C(N)CCCCn1nc(cc1)NC(=NCC(F)(F)F)N
Similarity: 0.579
Experimental value : Carcinogen
Predicted value : Carcinogen



Compound #5

CAS: 1582-09-8
Dataset id:775 (Training Set)
SMILES: O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F
Similarity: 0.579
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 70052-12-9
Dataset id:245 (Training Set)
SMILES: O=C(O)C(N)(CCCN)C(F)F
Similarity: 0.578
Experimental value : NON-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 = 0.632 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.531 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
⚠	Concordance for similar molecules Concordance index = 0.531 Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Model's descriptors range check Descriptors range check = False Explanation: 1descriptor(s) for this compound have values outside 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)..
✔	Model class assignment reliability Pos/Non-Pos difference = 0.473 Explanation: model class assignment is well defined..
✖	Neural map neurons concordance Neurons concordance = 0.5 Explanation: predicted substance falls into a neuron that is populated by no 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.

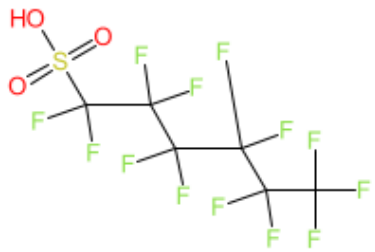
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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: OS 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)S The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

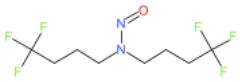


Prediction: Reliability:

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
- 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 found)
The following alerts have been found: SA43 Perfluorooctanoic acid (PFOA)

Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogen activity: Carcinogen
Structural Alerts: SA43 Perfluorooctanoic acid (PFOA)
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

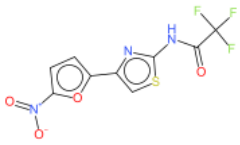
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 83335-32-4
Dataset id:496 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.654
Experimental value : Carcinogen
Predicted value : Carcinogen

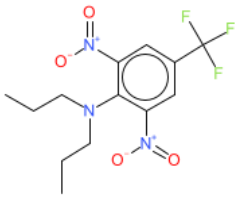
Alerts (not found also in the target): SA21 Alkyl and aryl N-nitroso groups



Compound #2

CAS: 42011-48-3
Dataset id:763 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.587
Experimental value : Carcinogen
Predicted value : Carcinogen

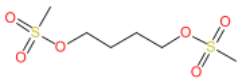
Alerts (not found also in the target): SA27 Nitro aromatic



Compound #3

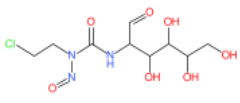
CAS: 1582-09-8
Dataset id:234 (Training Set)
SMILES: O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F
Similarity: 0.579
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): SA27 Nitro aromatic



Compound #4

CAS: 55-98-1
Dataset id:249 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.577
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

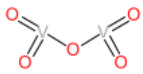


Compound #5

CAS: 54749-90-5
Dataset id:826 (Training Set)
SMILES: O=NN(C(=O)NC(C=O)C(O)C(O)C(O)CO)CCCl
Similarity: 0.576
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): SA8 Aliphatic halogens; SA11 Simple aldehyde; SA21 Alkyl and aryl N-nitroso groups

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



Compound #6

CAS: 1314-62-1
Dataset id:771 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.566
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.314 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.618 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..
✖	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 (2 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 0) Reasoning on fragments/structural alerts :.

Fragment found: SA43 Perfluorooctanoic acid (PFOA)

Perfluorooctanoic acid (PFOA)

No compounds with the same fragment have been found int the model's dataset.

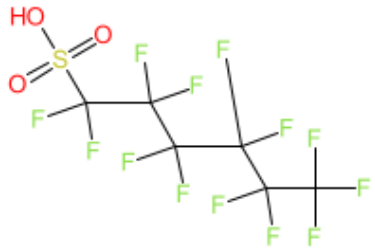
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

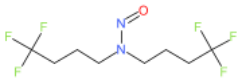


Prediction: Reliability:

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
- 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 (2 unknown fragments found)

Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)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

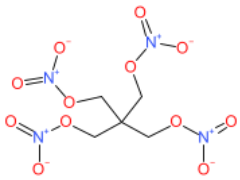
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

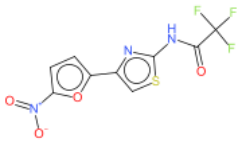
CAS: 83335-32-4
Dataset id:409 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.654
Experimental value : Carcinogen
Predicted value : Carcinogen

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



Compound #2

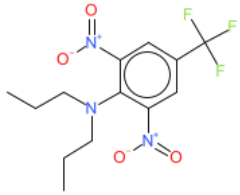
CAS: 78-11-5
Dataset id:971 (Training Set)
SMILES: O=[N+]([O-])OCC(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.591
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #3

CAS: 42011-48-3
Dataset id:596 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.587
Experimental value : Carcinogen
Predicted value : Carcinogen

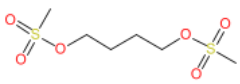
Alerts (not found also in the target): Carcinogenity alert no. 2; Carcinogenity alert no. 19



Compound #4

CAS: 1582-09-8
Dataset id:192 (Training Set)
SMILES: O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F
Similarity: 0.579
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 35; Carcinogenity alert no. 36; Carcinogenity alert no. 41; Carcinogenity alert no. 42

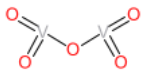


Compound #5

CAS: 55-98-1
Dataset id:662 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.577
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 8

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: 0.566
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.233 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.608 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 = 0.313 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.4 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 (2 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.

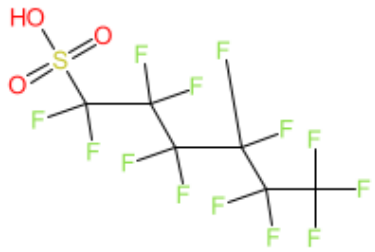
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

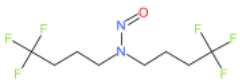


Prediction: Reliability:

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 optimal
- some 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 (2 unknown fragments and 1 infrequent_fragments found)
The following relevant fragments have been found: Carcinogenity alert no. 107; Carcinogenity alert no. 125

Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogenic activity: Carcinogen
No. alerts for carcinogenicity: 2
Structural Alerts: Carcinogenity alert no. 107; 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: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.654
Experimental value : Carcinogen
Predicted value : Carcinogen

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

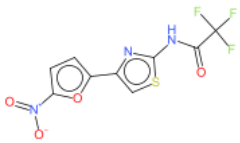
Alerts (not found also in the target): Carcinogenity alert no. 5; Carcinogenity alert no. 8; Carcinogenity alert no. 10; Carcinogenity alert no. 15; Carcinogenity alert no. 50; Carcinogenity alert no. 51; Carcinogenity alert no. 53; Carcinogenity alert no. 54; Carcinogenity alert no. 55; Carcinogenity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

Compound #2

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.612
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

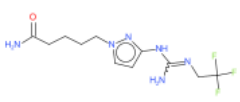


Compound #3

CAS: 42011-48-3
Dataset id:774 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.587
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. 63; Carcinogenity alert no. 64; Carcinogenity alert no. 90; Carcinogenity alert no. 108; Carcinogenity alert no. 117; Carcinogenity alert no. 123

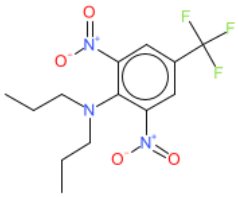


Compound #4

CAS: 84545-30-2
Dataset id:392 (Training Set)
SMILES: O=C(N)CCCCn1nc(cc1)NC(=NCC(F)(F)F)N
Similarity: 0.579
Experimental value : Carcinogen
Predicted value : Carcinogen

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

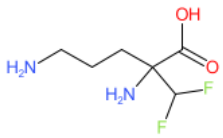
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #5

CAS: 1582-09-8
Dataset id:775 (Training Set)
SMILES: O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F
Similarity: 0.579
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 31; Carcinogenity alert no. 33; Carcinogenity alert no. 63; Carcinogenity alert no. 64



Compound #6

CAS: 70052-12-9
Dataset id:245 (Training Set)
SMILES: O=C(O)C(N)(CCCN)C(F)F
Similarity: 0.578
Experimental value : NON-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 = 0.219 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.615 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.672 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
⚠	Concordance for similar molecules Concordance index = 0.672 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 = 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.

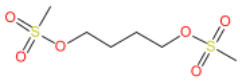
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on fragments/structural alerts - 1 of 2:.

Fragment found: Carcinogenity alert no. 107

Structural alert for carcinogenity defined by the SMARTS: C[S]=O

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

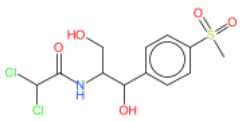


CAS: 55-98-1
Dataset id:490 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.577

Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 107

Alerts (not found also in the target): Carcinogenity alert no. 68

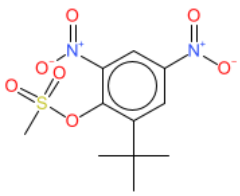


CAS: 15318-45-3
Dataset id:740 (Training Set)
SMILES: O=C(NC(CO)C(O)c1ccc(cc1)S(=O)(=O)C)C(Cl)Cl
Similarity: 0.544

Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 107

Alerts (not found also in the target): Carcinogenity alert no. 57; Carcinogenity alert no. 72



CAS: 29110-68-7
Dataset id:279 (Training Set)
SMILES: O=[N+]([O-])c1cc(c(OS(=O)(=O)C)c(c1)C(C)(C)C)[N+](=O)[O-]
Similarity: 0.53

Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 107

Alerts (not found also in the target): Carcinogenity alert no. 33; Carcinogenity alert no. 37; Carcinogenity alert no. 40; Carcinogenity alert no. 63; Carcinogenity alert no. 64

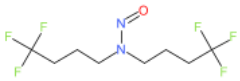
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on fragments/structural alerts - 2 of 2:.

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: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.654

Experimental value : Carcinogen
Predicted value : Carcinogen

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

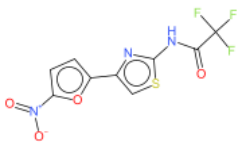
Alerts (not found also in the target): Carcinogenity alert no. 5; Carcinogenity alert no. 8; Carcinogenity alert no. 10; Carcinogenity alert no. 15; Carcinogenity alert no. 50; Carcinogenity alert no. 51; Carcinogenity alert no. 53; Carcinogenity alert no. 54; Carcinogenity alert no. 55; Carcinogenity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

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.612

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: 42011-48-3
Dataset id:774 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.587

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. 63; Carcinogenity alert no. 64; Carcinogenity alert no. 90; Carcinogenity alert no. 108; Carcinogenity alert no. 117; Carcinogenity alert no. 123

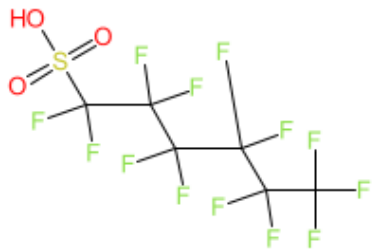
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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: OS 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)S The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

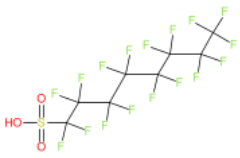


Prediction: Reliability:

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:
- 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)

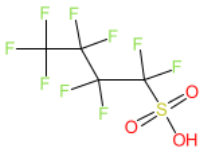
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Oral Carcinogenic class: NON-Carcinogen
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



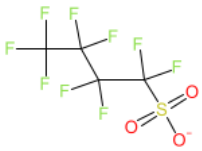
Compound #1

CAS: 1763-23-1
Dataset id:628 (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)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #2

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: 0.885
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

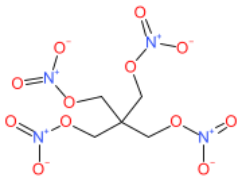


Compound #3

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: 0.875
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

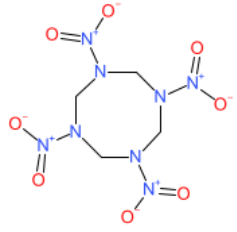
Compound #4

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: 0.819
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 78-11-5
Dataset id:253 (Test Set)
SMILES: O=[N+]([O-])OCC(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.591
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 2691-41-0
Dataset id:612 (Training Set)
SMILES: O=[N+]([O-])N1CN([N+](=O)[O-])CN([N+](=O)[O-])CN([N+](=O)[O-])C1
Similarity: 0.585
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✓	Global AD Index AD index = 0.807 Explanation: The predicted compound is into the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.902 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✓	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 = 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.

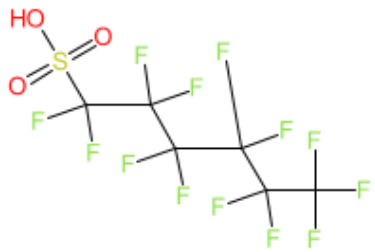
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S
The fragment has less than 3 occurrences in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

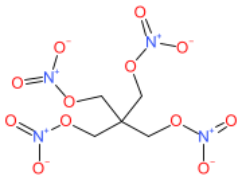


Prediction: Reliability:

Prediction is -1.53, 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 (3 unknown fragments and 1 infrequent_fragments found)

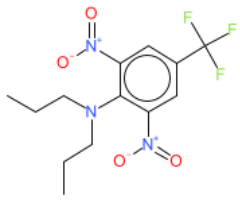
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Oral Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: -1.53
Predicted Oral Carcinogenicity SF [1/(mg/kg-day)]: 0.0293
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



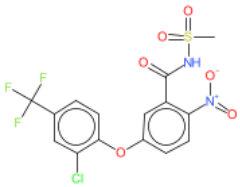
Compound #1

CAS: 78-11-5
Dataset id:253 (Training Set)
SMILES: O=[N+]([O-])OCC(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.591
Experimental value : -2.4
Predicted value : -2.004



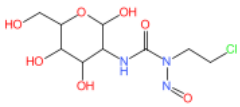
Compound #2

CAS: 1582-09-8
Dataset id:307 (Training Set)
SMILES: O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F
Similarity: 0.579
Experimental value : -2.11
Predicted value : -1.41



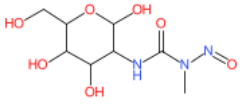
Compound #3

CAS: 72178-02-0
Dataset id:148 (Training Set)
SMILES: O=C(NS(=O)(=O)C)c2cc(Oc1ccc(cc1Cl)C(F)(F)F)ccc2[N+](=O)[O-]
Similarity: 0.575
Experimental value : -0.72
Predicted value : -1.975



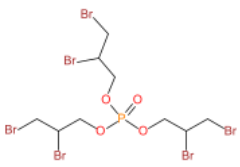
Compound #4

CAS: 54749-90-5
Dataset id:78 (Training Set)
SMILES: O=NN(C(=O)NC1C(O)OC(CO)C(O)C1(O))CCCl
Similarity: 0.566
Experimental value : 2.38
Predicted value : 2.294



Compound #5

CAS: 18883-66-4
Dataset id:278 (Test Set)
SMILES: O=NN(C(=O)NC1C(O)OC(CO)C(O)C1(O))C
Similarity: 0.56
Experimental value : 2.04
Predicted value : 2.315



Compound #6

CAS: 126-72-7
Dataset id:311 (Training Set)
SMILES: O=P(OCC(CBr)Br)(OCC(CBr)Br)OCC(CBr)Br
Similarity: 0.558
Experimental value : 0.36
Predicted value : 0.564

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.199 Explanation: The predicted compound is outside the Applicability Domain of the model.
✖	Similar molecules with known experimental value Similarity index = 0.585 Explanation: No similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.548 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✔	Concordance for similar molecules Concordance index = 0.722 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 = 0.7 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 (3 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.

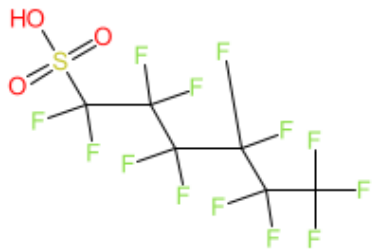
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S 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 0 -

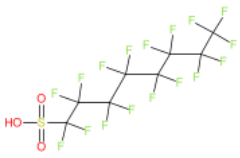


Prediction: Reliability:

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:
- similar molecules found in the training set have experimental values that disagree with the predicted value
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)

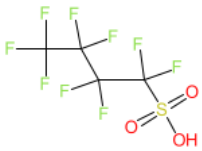
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Inhalation Carcinogenic class: Carcinogen
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



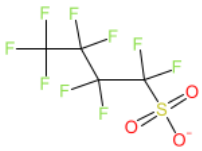
Compound #1

CAS: 1763-23-1
Dataset id:614 (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)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #2

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: 0.885
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

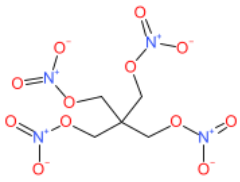


Compound #3

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: 0.875
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

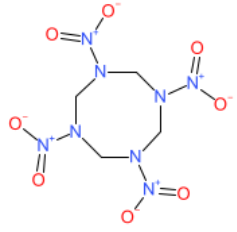
Compound #4

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: 0.819
Experimental value : NON-Carcinogen
Predicted value : Carcinogen



Compound #5

CAS: 78-11-5
Dataset id:611 (Training Set)
SMILES: O=[N+]([O-])OCC(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.591
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 2691-41-0
Dataset id:595 (Test Set)
SMILES: O=[N+]([O-])N1CN([N+](=O)[O-])CN([N+](=O)[O-])CN([N+](=O)[O-])C1
Similarity: 0.585
Experimental value : NON-Carcinogen
Predicted value : 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 = 0.902 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✔	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 0 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 = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 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.

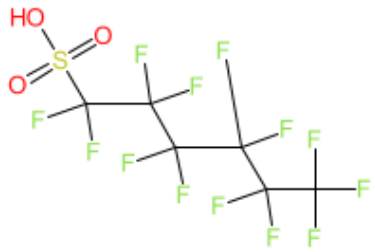
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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
	Fragment defined by the SMILES: CC(F)(F)S The fragment has less than 3 occurrences in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

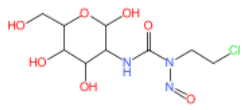


Prediction: Reliability:

Prediction is 0.14, 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 optimal
- 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 (4 unknown fragments found)

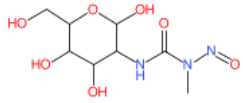
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Inhalation Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: 0.14
Predicted Inhalation Carcinogenicity SF [1/(mg/kg-day)]: 1.37
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



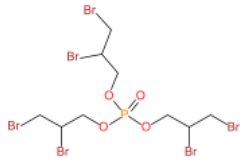
Compound #1

CAS: 54749-90-5
Dataset id:64 (Training Set)
SMILES: O=NN(C(=O)NC1C(O)OC(CO)C(O)C1(O))CCCl
Similarity: 0.566
Experimental value : 2.38
Predicted value : 1.702



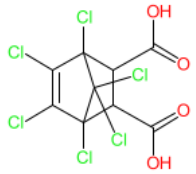
Compound #2

CAS: 18883-66-4
Dataset id:234 (Test Set)
SMILES: O=NN(C(=O)NC1C(O)OC(CO)C(O)C1(O))C
Similarity: 0.56
Experimental value : 2.04
Predicted value : 0.989



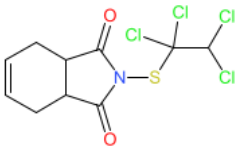
Compound #3

CAS: 126-72-7
Dataset id:257 (Training Set)
SMILES: O=P(OCC(CBr)Br)(OCC(CBr)Br)OCC(CBr)Br
Similarity: 0.558
Experimental value : 0.36
Predicted value : 0.569



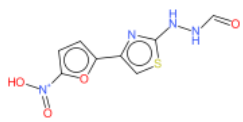
Compound #4

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: 0.543
Experimental value : -1.04
Predicted value : 1.623



Compound #5

CAS: 2425-06-1
Dataset id:47 (Training Set)
SMILES: O=C1N(C(=O)C2CC=CCC12)SC(C(Cl)Cl)(Cl)Cl
Similarity: 0.532
Experimental value : -0.82
Predicted value : -0.737



Compound #6

CAS: 3570-75-0
Dataset id:245 (Test Set)
SMILES: O=CNNc1nc(cs1)c2oc(cc2)[N+](=O)O
Similarity: 0.529
Experimental value : 0.36
Predicted value : -0.897

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.225 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.563 Explanation: No similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.865 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✗	Concordance for similar molecules Concordance index = 2.072 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 = 1.051 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 = 0.4 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 (4 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.

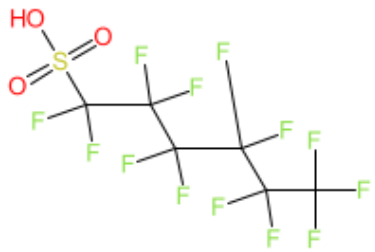
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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)S 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
	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 0 -

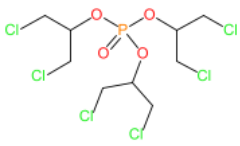


Prediction: Reliability:

Prediction is 4.1918, 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
- 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 (4 unknown fragments and 2 infrequent_fragments found)

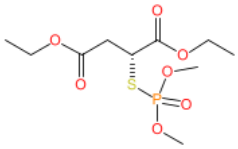
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: 4.1918
Predicted TD50 [mg/kg bw/day]: 0.0001
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



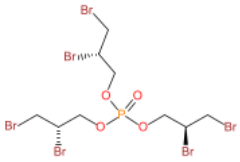
Compound #1

CAS: N.A.
Dataset id:18 (Training Set)
SMILES: P(=O)(OC(CCl)CCl)(OC(CCl)CCl)OC(CCl)CCl
Similarity: 0.562
Experimental value : -2.158
Predicted value : -2.535



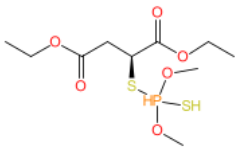
Compound #2

CAS: N.A.
Dataset id:38 (Training Set)
SMILES: CCOC(=O)[C@@H](CC(=O)OCC)SP(=O)(OC)OC
Similarity: 0.562
Experimental value : -3.193
Predicted value : -3.699



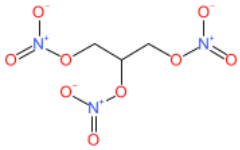
Compound #3

CAS: N.A.
Dataset id:12 (Training Set)
SMILES: P(=O)(OC[C@@H](CBr)Br)(OC[C@@H](CBr)Br)OC[C@@H](CBr)Br
Similarity: 0.558
Experimental value : -1.535
Predicted value : -1.665



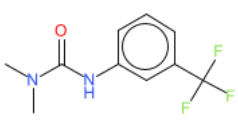
Compound #4

CAS: N.A.
Dataset id:26 (Training Set)
SMILES: P(S)(S[C@H](C(=O)OCC)CC(=O)OCC)(OC)OC
Similarity: 0.553
Experimental value : -2.614
Predicted value : -3.531



Compound #5

CAS: N.A.
Dataset id:24 (Training Set)
SMILES: C(C(CON(=O)=O)ON(=O)=O)ON(=O)=O
Similarity: 0.549
Experimental value : -2.344
Predicted value : -0.701



Compound #6

CAS: N.A.
Dataset id:87 (Training Set)
SMILES: c1(cc(ccc1)NC(=O)N(C)C)C(F)(F)F
Similarity: 0.547
Experimental value : -1.744
Predicted value : -0.804

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.191 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.562 Explanation: No similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.442 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✗	Concordance for similar molecules Concordance index = 6.867 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 = 0.506 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 (4 unknown fragments and 2 infrequent_fragments found)..

Symbols explanation:

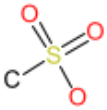
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

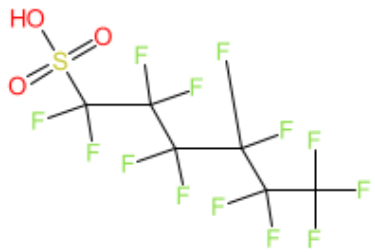
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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: OS 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)S The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CS(O)(=O)=O 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 0 -

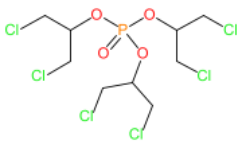


Prediction: Reliability:

Prediction is -6.1288, 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
- 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 (5 unknown fragments and 2 infrequent_fragments found)

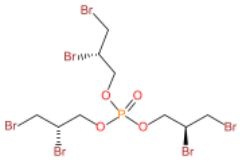
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: -6.1288
Predicted TD50 [mg/kg bw/day]: 1345259.17
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



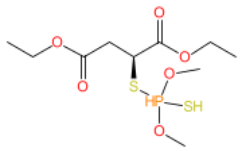
Compound #1

CAS: N.A.
Dataset id:92 (Training Set)
SMILES: P(=O)(OC(CCl)CCl)(OC(CCl)CCl)OC(CCl)CCl
Similarity: 0.562
Experimental value : -2.314
Predicted value : -1.883



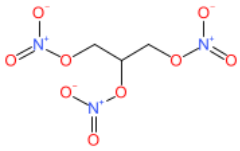
Compound #2

CAS: N.A.
Dataset id:106 (Training Set)
SMILES: P(=O)(OC[C@@H](CBr)Br)(OC[C@@H](CBr)Br)OC[C@@H](CBr)Br
Similarity: 0.558
Experimental value : -3.45
Predicted value : -3.074



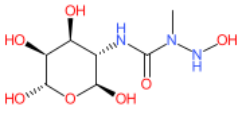
Compound #3

CAS: N.A.
Dataset id:107 (Training Set)
SMILES: P(S)(S[C@H](C(=O)OCC)CC(=O)OCC)(OC)OC
Similarity: 0.553
Experimental value : -3.588
Predicted value : -2.078



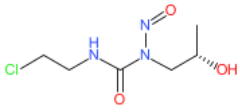
Compound #4

CAS: N.A.
Dataset id:94 (Training Set)
SMILES: C(C(CON(=O)=O)ON(=O)=O)ON(=O)=O
Similarity: 0.549
Experimental value : -2.517
Predicted value : -6.738



Compound #5

CAS: N.A.
Dataset id:8 (Training Set)
SMILES: [C@H]1([C@H]([C@@H]([C@@H]([C@H](O1)O)O)O)O)NC(=O)N(NO)C)O
Similarity: 0.546
Experimental value : -0.646
Predicted value : 3.177



Compound #6

CAS: N.A.
Dataset id:152 (Test Set)
SMILES: N(C[C@@H](O)C)(C(=O)NCCCl)N=O
Similarity: 0.539
Experimental value : 0.053
Predicted value : -0.968

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.19 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.56 Explanation: No similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.403 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✗	Concordance for similar molecules Concordance index = 3.247 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 = 0.431 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 (5 unknown fragments and 2 infrequent_fragments found)..

Symbols explanation:

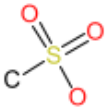
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

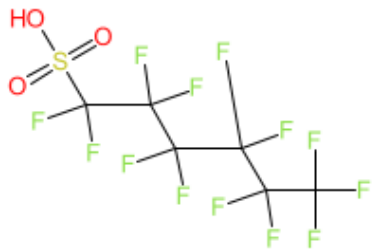
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) 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: O=S 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)S The fragment has never been found in the model's training set
	Fragment defined by the SMILES: CS(O)(=O)=O The fragment has never been found in the model's training set
	Fragment defined by the SMILES: OS 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 0 -

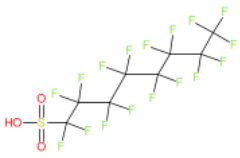


Prediction: Reliability:

Prediction is 203.28 mg/kg, the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections.

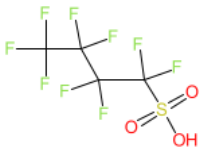
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted log LD50 [log(mmol/Kg)]: -0.294
Predicted log LD50 [mg/Kg]: 203.28
Molecules used for prediction: 3
Experimental value [mg/Kg]: -
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



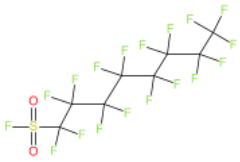
Compound #1

CAS: N.A.
Dataset id:4691 (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)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922
Experimental value : -0.52
Predicted value : -0.203



Compound #2

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: 0.885
Experimental value : 0.16
Predicted value : -



Compound #3

CAS: N.A.
Dataset id:4286 (Training Set)
SMILES: O=S(=O)(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: 0.861
Experimental value : -0.51
Predicted value : -0.478

Compound #4

CAS: N.A.
Dataset id:4690 (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: 0.819
Experimental value : 0
Predicted value : -0.017

Compound #5

CAS: N.A.
Dataset id:354 (Training Set)
SMILES: O=S(=O)(NCC)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: 0.814
Experimental value : -0.02
Predicted value : -0.324

Compound #6

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: 0.811
Experimental value : 0.62
Predicted value : 0.397

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 = 0.886 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.174 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0.299 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 = 0.317 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:

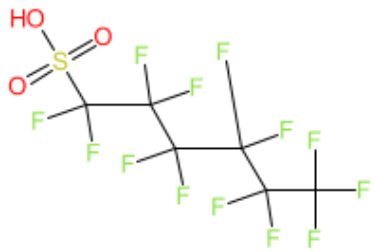
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 0 -



EXPERIMENTAL DATA

E xperimental value is 3.6 log(L/kg). Model prediction is 1.58 log(L/kg) (LOW reliability).
Warning: the prediction may be not fully reliable due to the presence of one or more fragments related to model outliers.
The following relevant fragments have been found: 10 F atoms in the molecule (SO 10); SO3H group (PG 02)

Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: 3.6
Predicted BCF [log(L/kg)]: 1.58
Predicted BCF [L/kg]: 38
Predicted BCF from sub-model 1 (HM) [log(L/kg)]: 1.55
Predicted BCF from sub-model 2 (GA) [log(L/kg)]: 1.86
Predicted LogP (MLogP): 2.85
Structural Alerts: 10 F atoms in the molecule (SO 10); SO3H group (PG 02)
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

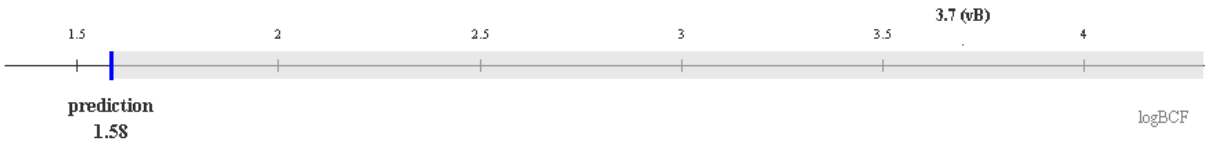
2. Possible Use and Uncertainty

Threshold 3.3 (bioaccumulative)

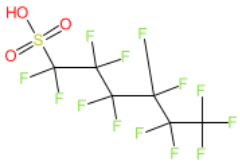
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.3, the current compound can not be associated (due to its Applicability Domain index value) to any conservative interval.
No safe classification can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



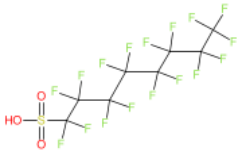
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 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)C(F)(F)F
Similarity: 1
Experimental value : 3.6
Predicted value : 1.585

Alerts (found also in the target): 10 F atoms in the molecule (SO 10); SO3H group (PG 02)



Compound #2

CAS: 1763-23-1
Dataset id:57 (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)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922
Experimental value : 3.73
Predicted value : 1.697

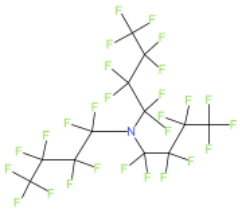
Alerts (found also in the target): 10 F atoms in the molecule (SO 10); SO3H group (PG 02)

Compound #3

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: 0.819
Experimental value : 3.12
Predicted value : 2.534

Alerts (found also in the target): 10 F atoms in the molecule (SO 10)

Alerts (not found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

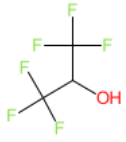


Compound #4

CAS: 311-89-7
Dataset id:419 (Training Set)
SMILES:
FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)N(C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.665
Experimental value : 1.42
Predicted value : 0.879

Alerts (found also in the target): 10 F atoms in the molecule (SO 10)

Alerts (not found also in the target): Tertiary amine (SR 05)

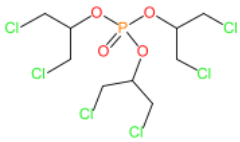


Compound #5

CAS: 920-66-1
Dataset id:263 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.638
Experimental value : 0.3
Predicted value : 0.601

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

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



Compound #6

CAS: 13674-87-8
Dataset id:400 (Test Set)
SMILES: O=P(OC(CCl)CCl)(OC(CCl)CCl)OC(CCl)CCl
Similarity: 0.562
Experimental value : 0.13
Predicted value : 1.777

Alerts (not found also in the target): 6 Cl atoms in the molecule (SO 01); PO2 residue (SR 03)

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 = 1 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✗	Accuracy of prediction for similar molecules Accuracy index = 2.015 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✗	Concordance for similar molecules Concordance index = 2.015 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 = 2.015 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 = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 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.

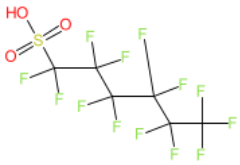
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on fragments/structural alerts - 1 of 2:.

Fragment found: 10 F atoms in the molecule (SO 10)

Compounds with ten or more F atoms fall into a chemical category that results out of the applicability domain of the model.

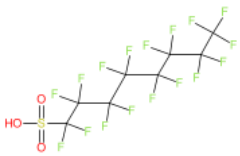
Following, the most similar compounds from the model's dataset having the same fragment.



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)C(F)(F)F
Similarity: 1

Experimental value : 3.6
Predicted value : 1.585

Alerts (found also in the target): 10 F atoms in the molecule (SO 10); SO3H group (PG 02)



CAS: 1763-23-1
Dataset id:57 (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)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922

Experimental value : 3.73
Predicted value : 1.697

Alerts (found also in the target): 10 F atoms in the molecule (SO 10); SO3H group (PG 02)

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: 0.819

Experimental value : 3.12
Predicted value : 2.534

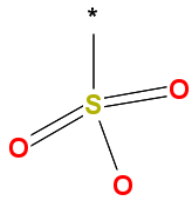
Alerts (found also in the target): 10 F atoms in the molecule (SO 10)

Alerts (not found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

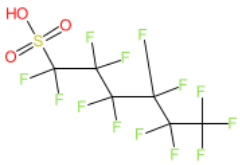
(Molecule 0) Reasoning on fragments/structural alerts - 2 of 2:.

Fragment found: SO3H group (PG 02)



This chemical contains a SO3H polar group. The presence of polar groups increases hydrophilicity, related to lower values of BCF.

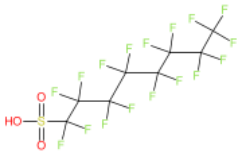
Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 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)C(F)(F)F
Similarity: 1

Experimental value : 3.6
Predicted value : 1.585

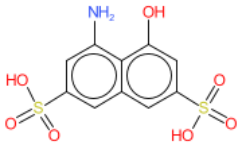
Alerts (found also in the target): 10 F atoms in the molecule (SO 10); SO3H group (PG 02)



CAS: 1763-23-1
Dataset id:57 (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)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922

Experimental value : 3.73
Predicted value : 1.697

Alerts (found also in the target): 10 F atoms in the molecule (SO 10); SO3H group (PG 02)



CAS: 90-20-0
Dataset id:493 (Test Set)
SMILES: O=S(=O)(O)c1cc(O)c2c(N)cc(cc2(c1))S(=O)(=O)O
Similarity: 0.55

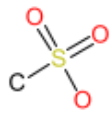
Experimental value : 0.46
Predicted value : 0.29

Alerts (found also in the target): SO3H group (PG 02)

Alerts (not found also in the target): Thiobenzene residue (SR 04)

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

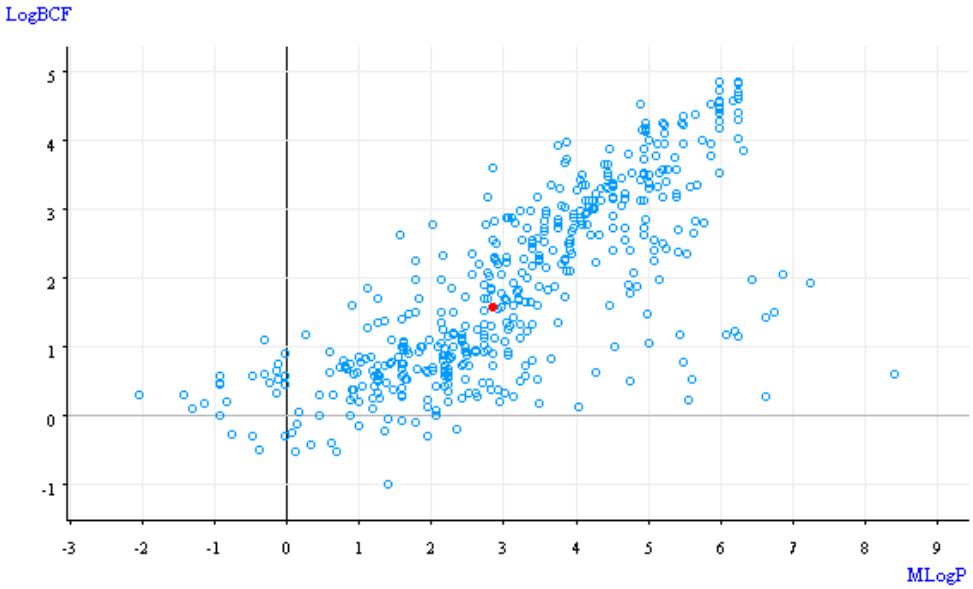
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

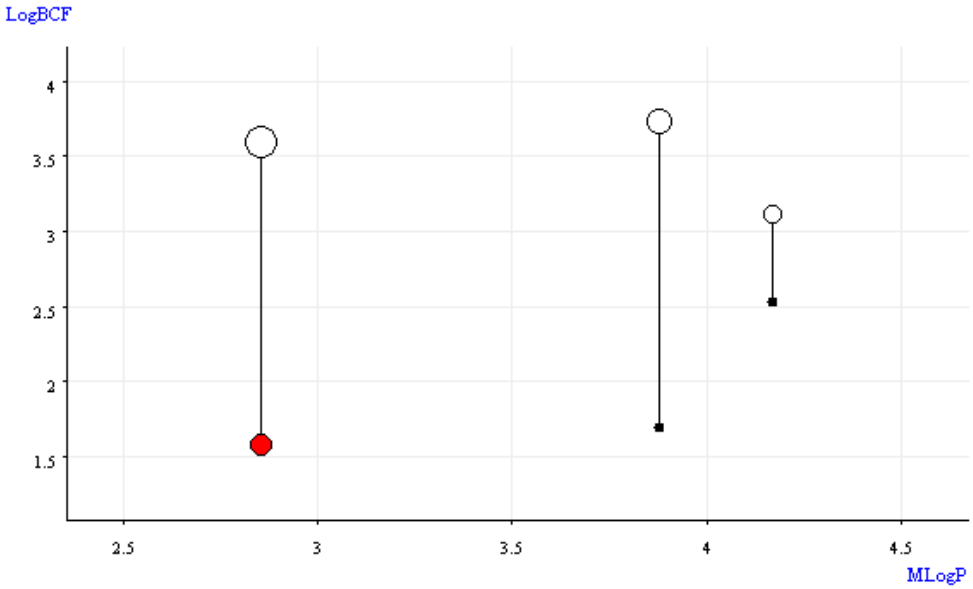
4.2 Reasoning:
Analysis of Molecular Descriptors

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.

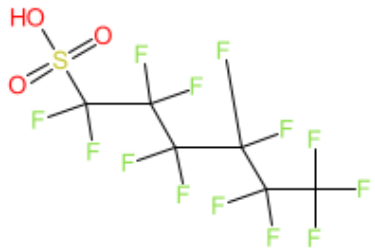


Following, a scatterplot of MLogP against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound;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



1. Prediction Summary

Prediction for compound Molecule 0 -

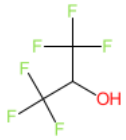


Prediction: Reliability:

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:
- 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
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
- reliability of logP value used by the model is not adequate
- 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 2 infrequent_fragments found)

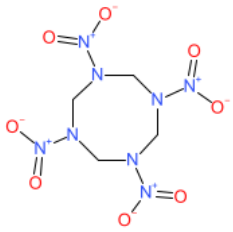
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF [log(L/kg)]: 0.5
Predicted BCF [L/kg]: 3
Predicted LogP (Meylan/Kowwin): 4.34
Predicted LogP reliability: Low
MW: 399.73
Ionic compound: yes
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



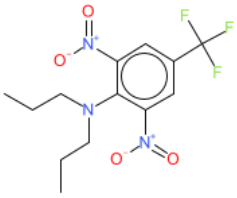
Compound #1

CAS: 920-66-1
Dataset id:117 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.638
Experimental value : 0.4
Predicted value : 0.762



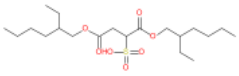
Compound #2

CAS: 2691-41-0
Dataset id:78 (Training Set)
SMILES: O=[N+]([O-])N1CN([N+](=O)[O-])CN([N+](=O)[O-])CN([N+](=O)[O-])C1
Similarity: 0.585
Experimental value : -0.3
Predicted value : 2.139



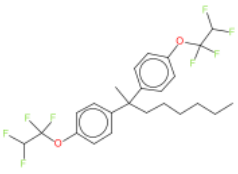
Compound #3

CAS: 1582-09-8
Dataset id:421 (Training Set)
SMILES: O=[N+]([O-])c1cc(cc(c1N(CCC)CCC)[N+](=O)[O-])C(F)(F)F
Similarity: 0.579
Experimental value : 3.36
Predicted value : 3.19



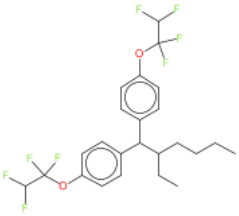
Compound #4

CAS: 577-11-7
Dataset id:28 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CC(C(=O)OCC(CC)CCCC)S(=O)(=O)O
Similarity: 0.571
Experimental value : 0.97
Predicted value : 0.5



Compound #5

CAS: N.A.
Dataset id:524 (Training Set)
SMILES: FC(F)C(F)(F)Oc1ccc(cc1)C(c2ccc(OC(F)(F)C(F)F)cc2)(C)CCCCCC
Similarity: 0.571
Experimental value : 2.85
Predicted value : 2.466



Compound #6

CAS: N.A.
Dataset id:523 (Training Set)
SMILES: FC(F)C(F)(F)Oc1ccc(cc1)C(c2ccc(OC(F)(F)C(F)F)cc2)C(CC)CCCC
Similarity: 0.569
Experimental value : 3.19
Predicted value : 2.483

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.207 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.61 Explanation: No similar compounds with known experimental value in the training set have been found..
✗	Accuracy of prediction for similar molecules Accuracy index = 1.401 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✓	Concordance for similar molecules Concordance index = 0.45 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 = 2.439 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 = 0 Explanation: reliability of logP value used by the model is not adequate..
✓	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 = 0.34 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 2 infrequent_fragments found)..

Symbols explanation:

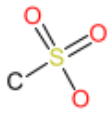
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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

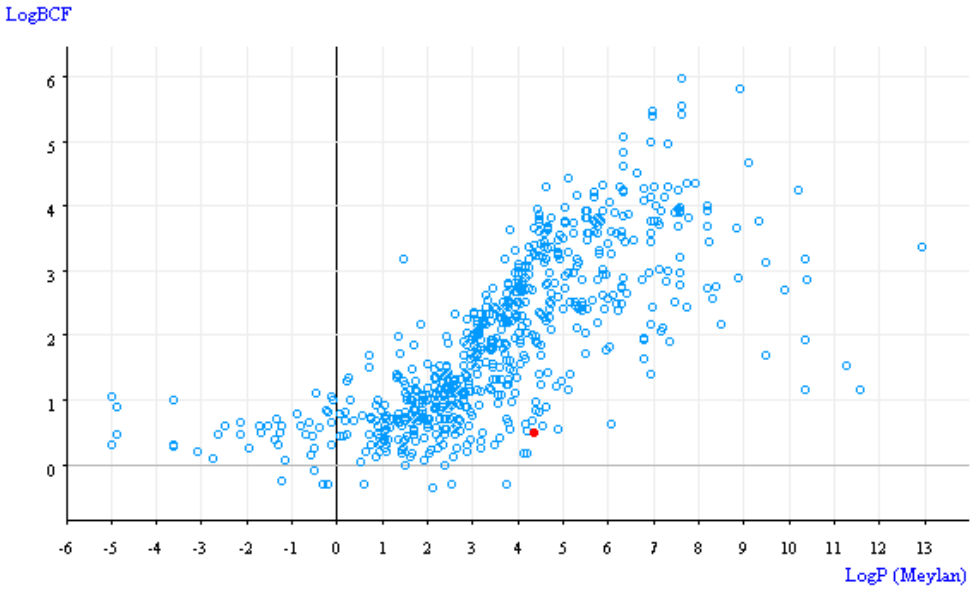
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

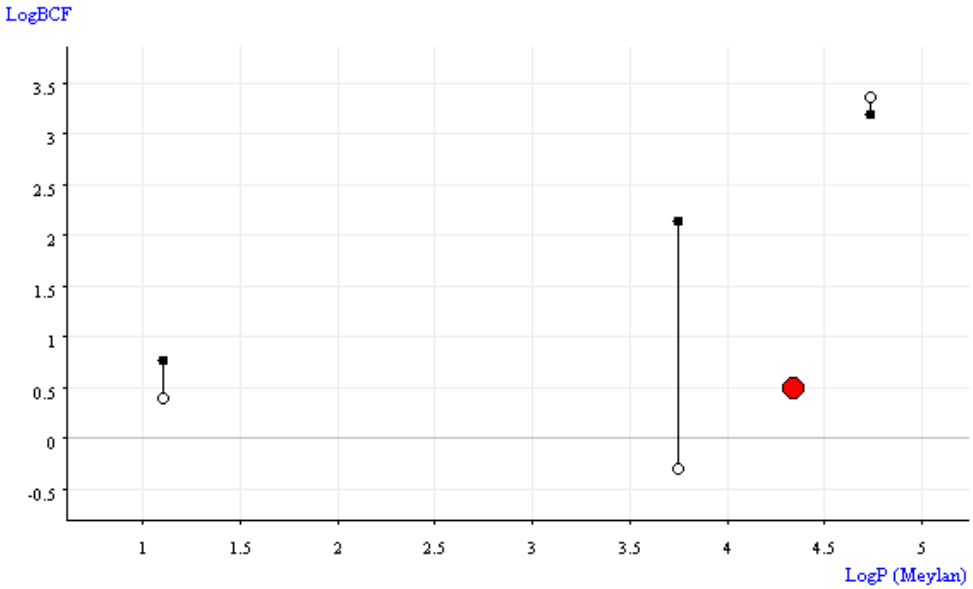
4.2 Reasoning:
Analysis of Molecular Descriptors

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.

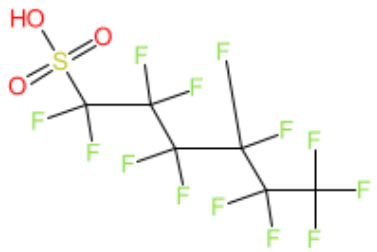


Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same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



1. Prediction Summary

Prediction for compound Molecule 0 -

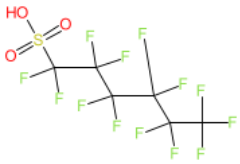


EXPERIMENTAL DATA

E xperimental value is 1.62 log(L/kg). Model prediction is 2.99 log(L/kg) (LOW reliability).

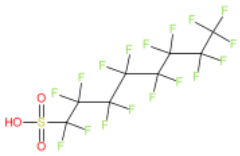
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: 1.62
Predicted BCF (up) [log(L/kg)]: 2.99
Predicted BCF (up) [L/kg]: 975
Predicted BCF (low) [log(L/kg)]: 2.94
Predicted BCF (low) [L/kg]: 868
Predicted BCF (mid) [log(L/kg)]: 2.96
Predicted BCF (mid) [L/kg]: 912
Predicted LogP (Meylan/Kowwin): 4.34
Predicted LogP reliability: Low
Predicted kM (Meylan): 0.62
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: 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)C(F)(F)F
Similarity: 1
Experimental value : 1.62
Predicted value : 2.989



Compound #2

CAS: 2795-39-3
Dataset id:79 (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)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922
Experimental value : 3.467
Predicted value : 3.79

Compound #3

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: 0.819
Experimental value : 0.977
Predicted value : 4

Compound #4

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: 0.769
Experimental value : 3.04
Predicted value : 3.236

Compound #5

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: 0.748
Experimental value : 3.72
Predicted value : 2.565

Compound #6

CAS: 307-55-1
Dataset id:413 (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)C(F)(F)F
Similarity: 0.728
Experimental value : 4.373
Predicted value : 1.714

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 = 1 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✗	Accuracy of prediction for similar molecules Accuracy index = 1.369 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✗	Concordance for similar molecules Concordance index = 1.369 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 = 1.369 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 = 0 Explanation: reliability of logP value used by the model is not adequate..
⚠	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

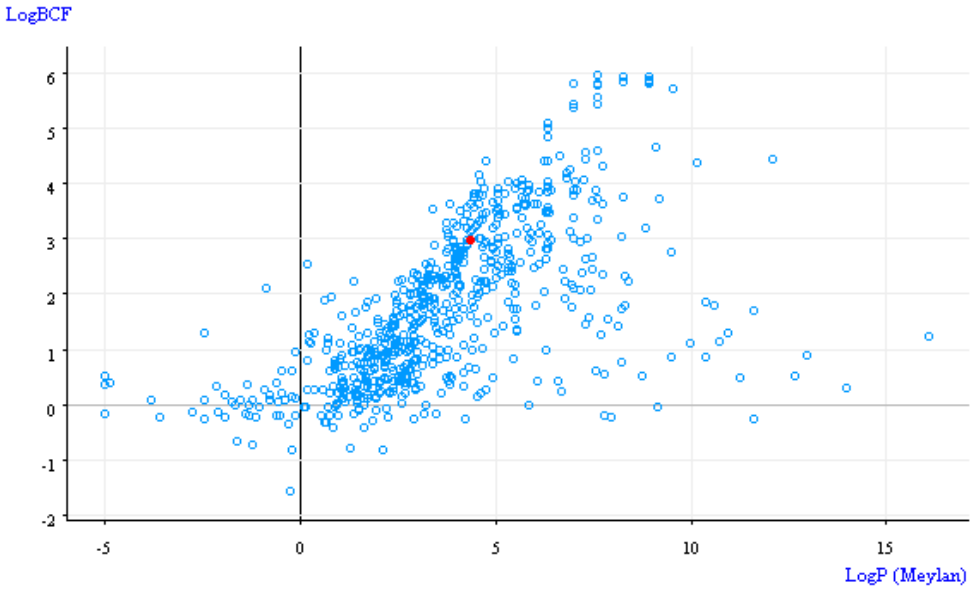
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

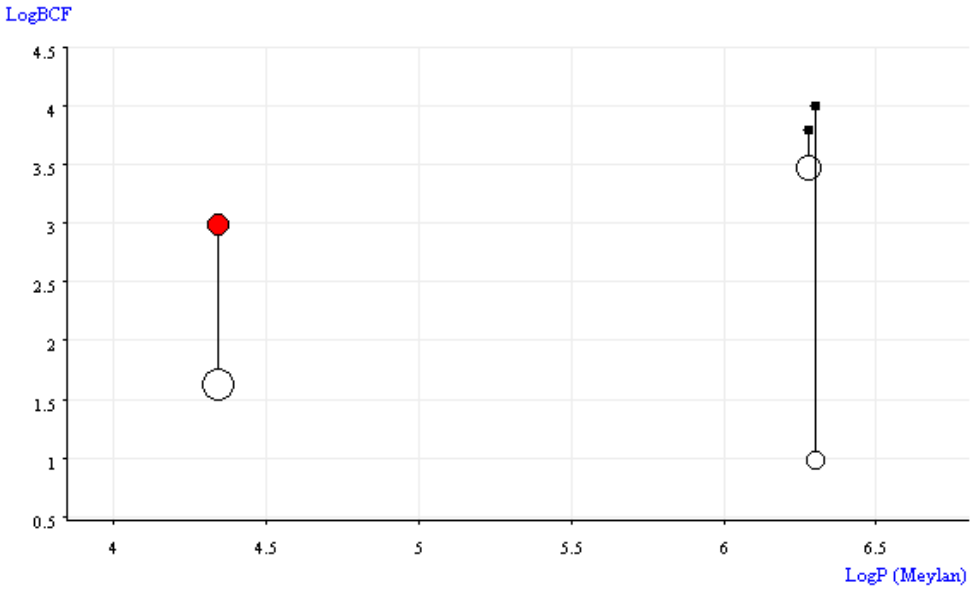
4.2 Reasoning:
Analysis of Molecular Descriptors

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.

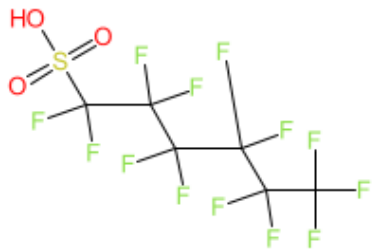


Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same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



1. Prediction Summary

Prediction for compound Molecule 0 -

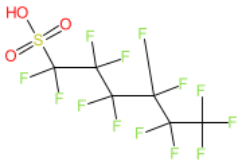


EXPERIMENTAL DATA

E xperimental value is 3.6 log(L/kg). Model prediction is 3.6 log(L/kg) (GOOD reliability).

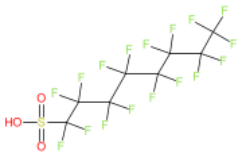
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: 3.6
Predicted BCF [log(L/kg)]: 3.6
Molecules used for prediction: 1
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



Compound #1

CAS: 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)C(F)(F)F
Similarity: 1
Experimental value : 3.6
Predicted value : 3.479

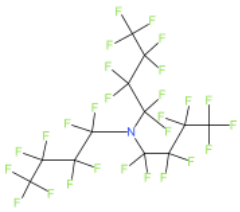


Compound #2

CAS: 1763-23-1
Dataset id:419 (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)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.922
Experimental value : 3.73
Predicted value : 2.948

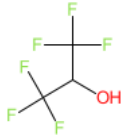
Compound #3

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: 0.819
Experimental value : 3.12
Predicted value : 3.096



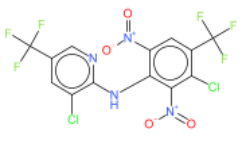
Compound #4

CAS: 311-89-7
Dataset id:303 (Training Set)
SMILES:
FC(F)(F)C(F)(F)C(F)(F)C(F)(F)N(C(F)(F)C(F)(F)C(F)(F)C(F)(F)F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.665
Experimental value : 1.299
Predicted value : 3.432



Compound #5

CAS: 920-66-1
Dataset id:391 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.638
Experimental value : 0.244
Predicted value : 1.753



Compound #6

CAS: 79622-59-6
Dataset id:774 (Training Set)
SMILES: O=[N+]([O-])c2cc(c(c(c2(Nc1ncc(cc1Cl)C(F)(F)F))[N+](=O)[O-])Cl)C(F)(F)F
Similarity: 0.609
Experimental value : 3.011
Predicted value : 3.136

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 = 0.121 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 = 0.121 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 = 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 0) 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)S
The fragment has less than 3 occurrences in the model's training set

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