VEGA QSAR MODELS

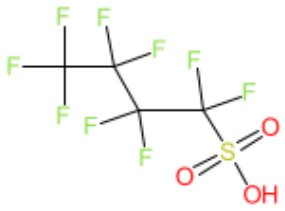
# Report

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 -



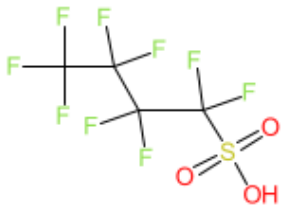

Prediction:

**Prediction is NON-Mutagenic with a consensus score of 0.15, 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)F
Used models: 4
Predicted Consensus Mutagen activity: NON-Mutagenic
Mutagenic Score: 0.05
Non-Mutagenic Score: 0.15
Model Caesar assessment: NON-Mutagenic (LOW reliability)
Model ISS assessment: NON-Mutagenic (LOW reliability)
Model SarPy assessment: NON-Mutagenic (LOW reliability)
Model KNN assessment: Mutagenic (LOW reliability)
Remarks:
none

# 1. Prediction Summary

Prediction for compound Molecule 0 -

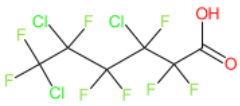



Prediction: Reliability:

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)

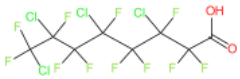
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



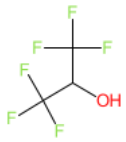
Compound #1

CAS: 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: 0.762
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #2

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: 0.722
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #3

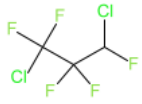
CAS: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.717
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #4

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: 0.683
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #5

CAS: 333-27-7
Dataset id:511 (Training Set)
SMILES: O=S(=O)(OC)C(F)(F)F
Similarity: 0.678
Experimental value : Mutagenic
Predicted value : Mutagenic



Compound #6

CAS: 507-55-1
Dataset id:34 (Training Set)
SMILES: FC(C(F)(F)C(F)(F)Cl)Cl
Similarity: 0.676
Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

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

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.513<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.732<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ✔ | Accuracy of prediction for similar molecules<br>Accuracy index = 1<br>Explanation: Accuracy of prediction for similar molecules found in the training set is good.. |
| ✔ | Concordance for similar molecules<br>Concordance index = 1<br>Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.. |
| ✔ | Model's descriptors range check<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.6<br>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.

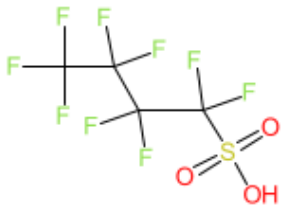
## 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

(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 -

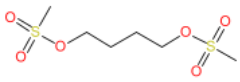



Prediction: Reliability:

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:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)

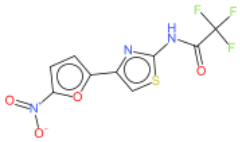
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



Compound #1

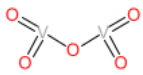
CAS: 55-98-1
Dataset id:249 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.622
Experimental value : Mutagenic
Predicted value : NON-Mutagenic



Compound #2

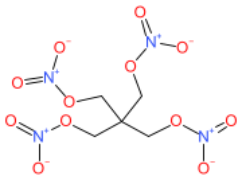
CAS: 42011-48-3
Dataset id:763 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.608
Experimental value : Mutagenic
Predicted value : Mutagenic

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



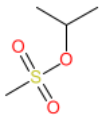
Compound #3

CAS: 1314-62-1
Dataset id:771 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.605
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

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: 0.601
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

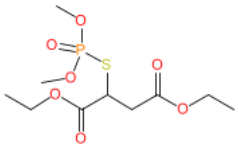


Compound #5

CAS: 926-06-7
Dataset id:287 (Training Set)
SMILES: O=S(=O)(OC(C)C)C
Similarity: 0.585
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA2 Alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid

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



Compound #6

CAS: 1634-78-2
Dataset id:118 (Training Set)
SMILES: O=C(OCC)CC(C(=O)OCC)SP(=O)(OC)OC
Similarity: 0.583
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA2 Alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.615<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ✖ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.489<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 0<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.34<br>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.

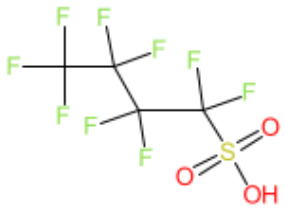
## 4.1 Reasoning:
### Relevant Chemical Fragments and Moieties

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

# 1. Prediction Summary

Prediction for compound Molecule 0 -

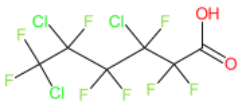



Prediction: Reliability:

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)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

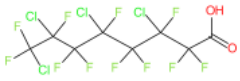


Compound #1

CAS: 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: 0.762
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

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

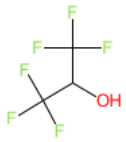


Compound #2

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: 0.722
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 #3

CAS: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.717
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM150

Compound #4

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: 0.683
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

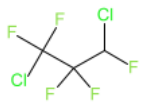
Alerts (found also in the target): SM150

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

Compound #5

CAS: 333-27-7
Dataset id:511 (Training Set)
SMILES: O=S(=O)(OC)C(F)(F)F
Similarity: 0.678
Experimental value : Mutagenic
Predicted value : Possible NON-Mutagenic

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




Compound #6

CAS: 507-55-1
Dataset id:34 (Training Set)
SMILES: FC(C(F)(F)C(F)(F)Cl)Cl
Similarity: 0.676
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

Alerts (not found also in the target): SM106

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✗ | Global AD Index<br>AD index = 0.387<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.732<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ✗ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.322<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ✓ | Concordance for similar molecules<br>Concordance index = 1<br>Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.. |
| ✗ | Atom Centered Fragments similarity check<br>ACF index = 0.6<br>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.

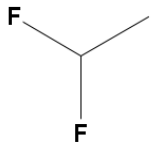
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

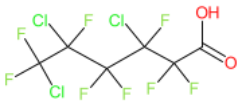
(Molecule 0) Reasoning on fragments/structural alerts - 1 of 2:.

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.

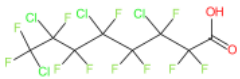


CAS: 2106-54-9
Dataset id:384 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl
Similarity: 0.762

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM150

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

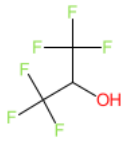


CAS: 2923-68-4
Dataset id:2832 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl)Cl
Similarity: 0.722

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: 920-66-1
Dataset id:3517 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.717

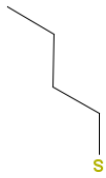
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM150

# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

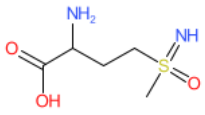
(Molecule 0) Reasoning on fragments/structural alerts - 2 of 2:.

Fragment found: SM153




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

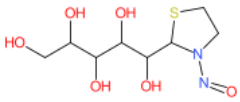
Following, the most similar compounds from the model's dataset having the same fragment.




CAS: 1982-67-8
Dataset id:309 (Training Set)
SMILES: O=C(O)C(N)CCS(=O)(=N)C
Similarity: 0.605

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

Alerts (found also in the target): SM153

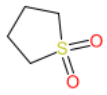



CAS: 92134-96-8
Dataset id:1424 (Training Set)
SMILES: O=NN1CCSC1C(O)C(O)C(O)C(O)CO
Similarity: 0.582

Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM153

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




CAS: 126-33-0
Dataset id:365 (Training Set)
SMILES: O=S1(=O)(CCCC1)
Similarity: 0.564

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

Alerts (found also in the target): SM153

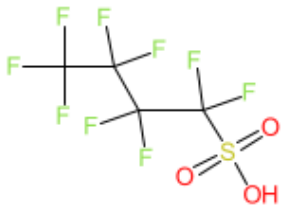
## 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

(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 -

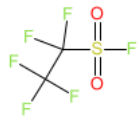



Prediction: Reliability:

Prediction is Mutagenic, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- some similar molecules found in the training set have experimental values that disagree with the predicted value

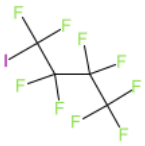
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: Mutagenic
Molecules used for prediction: 4
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



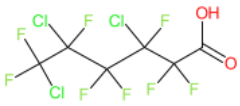
Compound #1

CAS: 354-87-0
Dataset id:5751 (Training Set)
SMILES: O=S(=O)(F)C(F)(F)C(F)(F)F
Similarity: 0.782
Experimental value : Mutagenic
Predicted value : Mutagenic



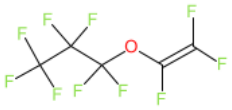
Compound #2

CAS: 423-39-2
Dataset id:3062 (Training Set)
SMILES: FC(F)(F)C(F)(F)C(F)(F)C(F)(F)I
Similarity: 0.768
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



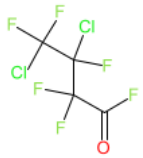
Compound #3

CAS: 2106-54-9
Dataset id:2157 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(C(F)(F)C(F)(C(F)(F)Cl)Cl)Cl
Similarity: 0.762
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



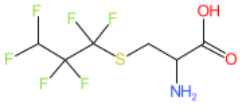
Compound #4

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: 0.751
Experimental value : Mutagenic
Predicted value : NON-Mutagenic



Compound #5

CAS: 678-06-8
Dataset id:5701 (Training Set)
SMILES: O=C(F)C(F)(F)C(F)(C(F)(F)Cl)Cl
Similarity: 0.749
Experimental value : Mutagenic
Predicted value : NON-Mutagenic



Compound #6

CAS: 98640-41-6
Dataset id:5515 (Training Set)
SMILES: O=C(O)C(N)CSC(F)(F)C(F)(F)C(F)F
Similarity: 0.74
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.62<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.765<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ⚠ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.506<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal.. |
| ⚠ | Concordance for similar molecules<br>Concordance index = 0.501<br>Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✔ | Atom Centered Fragments similarity check<br>ACF index = 1<br>Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.. |

Symbols explanation:

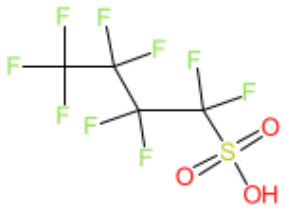
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 -

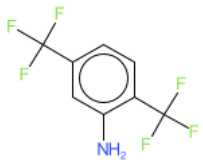



Prediction: Reliability:

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:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (3 unknown fragments found)

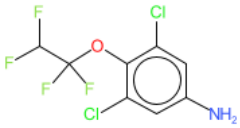
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Mutagen activity: NA
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



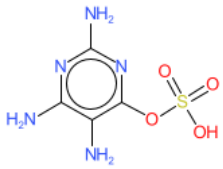
Compound #1

CAS: N.A.
Dataset id:7541 (Training Set)
SMILES: FC(F)(F)c1ccc(c(N)c1)C(F)(F)F
Similarity: 0.621
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



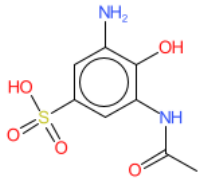
Compound #2

CAS: N.A.
Dataset id:5582 (Training Set)
SMILES: FC(F)C(F)(F)Oc1c(cc(N)cc1Cl)Cl
Similarity: 0.591
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



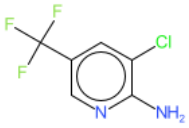
Compound #3

CAS: N.A.
Dataset id:7579 (Training Set)
SMILES: O=S(=O)(O)Oc1nc(nc(N)c1(N))N
Similarity: 0.589
Experimental value : Mutagenic
Predicted value : NON-Mutagenic



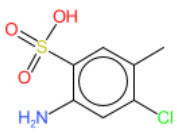
Compound #4

CAS: N.A.
Dataset id:4235 (Training Set)
SMILES: O=C(Nc1cc(cc(N)c1(O))S(=O)(=O)O)C
Similarity: 0.57
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



Compound #5

CAS: N.A.
Dataset id:4740 (Training Set)
SMILES: FC(F)(F)c1cnc(N)c(c1)Cl
Similarity: 0.566
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #6

CAS: N.A.
Dataset id:192 (Training Set)
SMILES: O=S(=O)(O)c1cc(c(cc1(N))Cl)C
Similarity: 0.561
Experimental value : Mutagenic
Predicted value : Mutagenic

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.605<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ✖ | Accuracy of prediction for similar molecules<br>Accuracy index = 0<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 0<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.4<br>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:

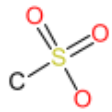
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

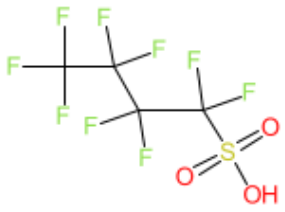
## 4.1 Reasoning:
### Relevant Chemical Fragments and Moieties

(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<br>The fragment has never been found in the model's training set |
|  | Fragment defined by the SMILES: CC(F)(F)S<br>The fragment has never been found in the model's training set |
|  | Fragment defined by the SMILES: CS(O)(=O)=O<br>The fragment has never been found in the model's training set |

# 1. Prediction Summary

Prediction for compound Molecule 0 -

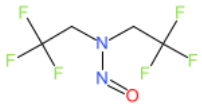



Prediction: Reliability:

Prediction is NON-Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- 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)

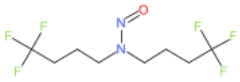
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Carcinogen activity: NON-Carcinogen
P(Carcinogen): 0.449
P(NON-Carcinogen): 0.551
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



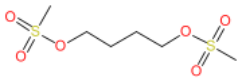
Compound #1

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.679
Experimental value : NON-Carcinogen
Predicted value : Carcinogen



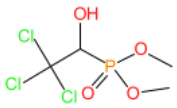
Compound #2

CAS: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.673
Experimental value : Carcinogen
Predicted value : Carcinogen



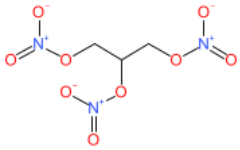
Compound #3

CAS: 55-98-1
Dataset id:490 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.622
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



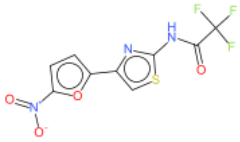
Compound #4

CAS: 52-68-6
Dataset id:770 (Test Set)
SMILES: O=P(OC)(OC)C(O)C(Cl)(Cl)Cl
Similarity: 0.609
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 55-63-0
Dataset id:782 (Training Set)
SMILES: O=[N+]([O-])OCC(O[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.609
Experimental value : Carcinogen
Predicted value : Carcinogen



Compound #6

CAS: 42011-48-3
Dataset id:774 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.608
Experimental value : Carcinogen
Predicted value : Carcinogen

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.676<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ✖ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.496<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ⚠ | Concordance for similar molecules<br>Concordance index = 0.504<br>Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Model's descriptors range check<br>Descriptors range check = False<br>Explanation: 1descriptor(s) for this compound have values outside the descriptor range of the compounds of the training set... |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.34<br>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<br>Pos/Non-Pos difference = 0.101<br>Explanation: model class assignment is well defined.. |
| ✔ | Neural map neurons concordance<br>Neurons concordance = 1<br>Explanation: predicted value agrees with experimental values of training set compounds laying in the same neuron.. |

Symbols explanation:

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

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

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

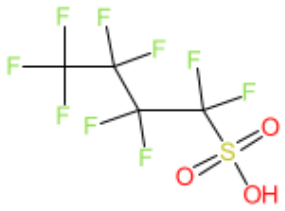
## 4.1 Reasoning:
### Relevant Chemical Fragments and Moieties

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

# 1. Prediction Summary

Prediction for compound Molecule 0 -

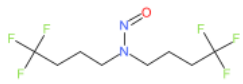



Prediction: Reliability:

Prediction is NON-Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (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)F
Experimental value: -
Predicted Carcinogen activity: NON-Carcinogen
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

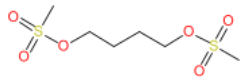
## 3.1 Applicability Domain:
### Similar Compounds, with Predicted and Experimental Values



Compound #1

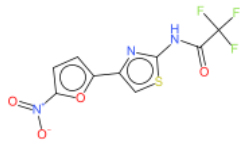
CAS: 83335-32-4
Dataset id:496 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.673
Experimental value : Carcinogen
Predicted value : Carcinogen

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



Compound #2

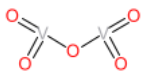
CAS: 55-98-1
Dataset id:249 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.622
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



Compound #3

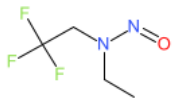
CAS: 42011-48-3
Dataset id:763 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.608
Experimental value : Carcinogen
Predicted value : Carcinogen

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



Compound #4

CAS: 1314-62-1
Dataset id:771 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.605
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

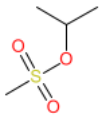


Compound #5

CAS: 82018-90-4
Dataset id:560 (Training Set)
SMILES: O=NN(CC)CC(F)(F)F
Similarity: 0.601
Experimental value : Carcinogen
Predicted value : Carcinogen

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

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



Compound #6

CAS: 926-06-7
Dataset id:287 (Training Set)
SMILES: O=S(=O)(OC(C)C)C
Similarity: 0.585
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): SA2 Alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.646<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ⚠ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.535<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 0<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.4<br>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.

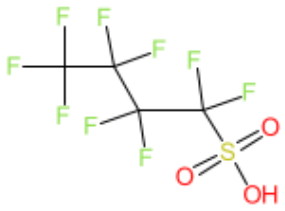
## 4.1 Reasoning:
### Relevant Chemical Fragments and Moieties

(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<br>The fragment has never been found in the model's training set |
|  | Fragment defined by the SMILES: CC(F)(F)S<br>The fragment has never been found in the model's training set |

# 1. Prediction Summary

Prediction for compound Molecule 0 -

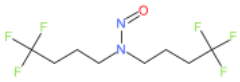



Prediction: Reliability:

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
- Accuracy of prediction for similar molecules found in the training set is not optimal
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (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)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

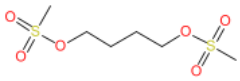
## 3.1 Applicability Domain:
### Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 83335-32-4
Dataset id:409 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.673
Experimental value : Carcinogen
Predicted value : Carcinogen

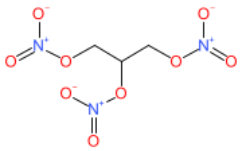
Alerts (not found also in the target): Carcinogenity alert no. 1; Carcinogenity alert no. 14; Carcinogenity alert no. 27



Compound #2

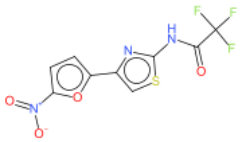
CAS: 55-98-1
Dataset id:662 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.622
Experimental value : Carcinogen
Predicted value : Carcinogen

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



Compound #3

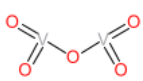
CAS: 55-63-0
Dataset id:917 (Training Set)
SMILES: O=[N+]([O-])OCC(O[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.609
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #4

CAS: 42011-48-3
Dataset id:596 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.608
Experimental value : Carcinogen
Predicted value : Carcinogen

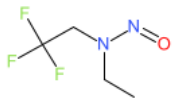
Alerts (not found also in the target): Carcinogenity alert no. 2; Carcinogenity alert no. 19



Compound #5

CAS: 1314-62-1
Dataset id:599 (Training Set)
SMILES: O=[V](=O)O[V](=O)=O
Similarity: 0.605
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen

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




Compound #6

CAS: 82018-90-4
Dataset id:470 (Training Set)
SMILES: O=NN(CC)CC(F)(F)F
Similarity: 0.601
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

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.632<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ⚠ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.691<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 0<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.4<br>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.

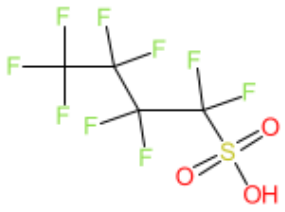
# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

(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<br>The fragment has never been found in the model's training set |
|  | Fragment defined by the SMILES: CC(F)(F)S<br>The fragment has never been found in the model's training set |

# 1. Prediction Summary

Prediction for compound Molecule 0 -

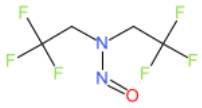



Prediction: Reliability:

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 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 (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)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

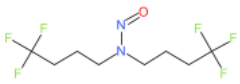


Compound #1

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.679
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

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

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

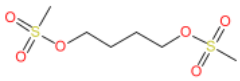


Compound #2

CAS: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.673
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

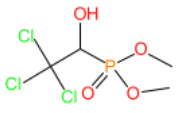


Compound #3

CAS: 55-98-1
Dataset id:490 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.622
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

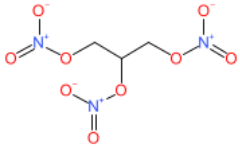


Compound #4

CAS: 52-68-6
Dataset id:770 (Test Set)
SMILES: O=P(OC)(OC)C(O)C(Cl)(Cl)Cl
Similarity: 0.609
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 57; Carcinogenity alert no. 98

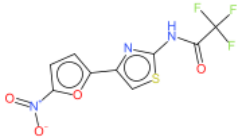
## 3.1 Applicability Domain:
### Similar Compounds, with Predicted and Experimental Values



Compound #5

CAS: 55-63-0
Dataset id:782 (Training Set)
SMILES: O=[N+]([O-])OCC(O[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.609
Experimental value : Carcinogen
Predicted value : Carcinogen

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



Compound #6

CAS: 42011-48-3
Dataset id:774 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.608
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

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.162<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.656<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ✖ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.347<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 0.347<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.34<br>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.

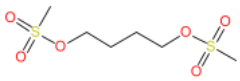
# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

(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.

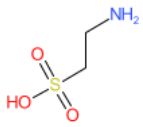


CAS: 55-98-1
Dataset id:490 (Training Set)
SMILES: O=S(=O)(OCCCCOS(=O)(=O)C)C
Similarity: 0.622

Experimental value : NON-Carcinogen
Predicted value : Carcinogen

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

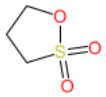
Alerts (not found also in the target): Carcinogenity alert no. 68



CAS: 107-35-7
Dataset id:718 (Training Set)
SMILES: O=S(=O)(O)CCN
Similarity: 0.598

Experimental value : NON-Carcinogen
Predicted value : Carcinogen

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



CAS: 1120-71-4
Dataset id:671 (Test Set)
SMILES: O=S1(=O)(OCCC1)
Similarity: 0.57

Experimental value : 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

## 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

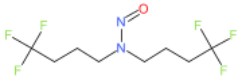
Following, the most similar compounds from the model's dataset having the same fragment.

CAS: 625-89-8
Dataset id:576 (Test Set)
SMILES: O=NN(CC(F)(F)F)CC(F)(F)F
Similarity: 0.679

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

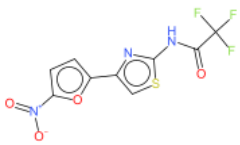


CAS: 83335-32-4
Dataset id:541 (Training Set)
SMILES: O=NN(CCCC(F)(F)F)CCCC(F)(F)F
Similarity: 0.673

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: 42011-48-3
Dataset id:774 (Training Set)
SMILES: O=C(Nc1nc(cs1)c2oc(cc2)[N+](=O)[O-])C(F)(F)F
Similarity: 0.608

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

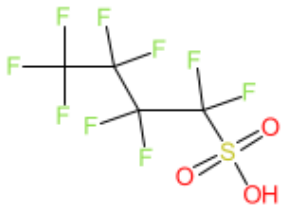
# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

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

# 1. Prediction Summary

Prediction for compound Molecule 0 -

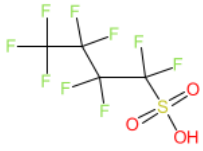



**EXPERIMENTAL DATA**

**E xperimental value is NON-Carcinogen. Model prediction is NON-Carcinogen (GOOD reliability).**

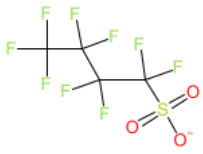
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: NON-Carcinogen
Predicted Oral Carcinogenic class: NON-Carcinogen
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



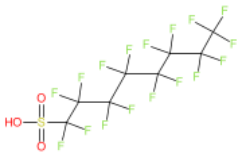
Compound #1

CAS: 375-73-5
Dataset id:627 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 1
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #2

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: 0.983
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

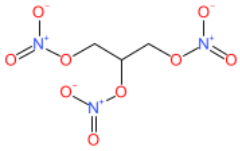


Compound #3

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)C(F)(F)F
Similarity: 0.809
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

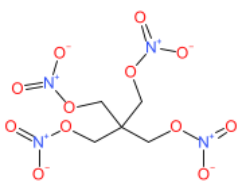
Compound #4

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)C(F)(F)F
Similarity: 0.732
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 55-63-0
Dataset id:218 (Training Set)
SMILES: O=[N+]([O-])OCC(O[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.609
Experimental value : Carcinogen
Predicted value : Carcinogen



Compound #6

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: 0.601
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✓ | Global AD Index<br>AD index = 0.85<br>Explanation: The predicted compound is into the Applicability Domain of the model. |
| ✓ | Similar molecules with known experimental value<br>Similarity index = 1<br>Explanation: Strongly similar compounds with known experimental value in the training set have been .. |
| ✓ | Accuracy of prediction for similar molecules<br>Accuracy index = 1<br>Explanation: Accuracy of prediction for similar molecules found in the training set is good.. |
| ✓ | Concordance for similar molecules<br>Concordance index = 1<br>Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.. |
| ✓ | Model's descriptors range check<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ⚠ | Atom Centered Fragments similarity check<br>ACF index = 0.85<br>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.

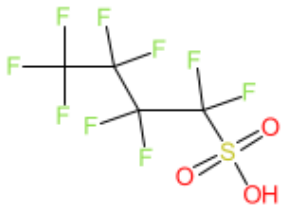
## 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

(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 -

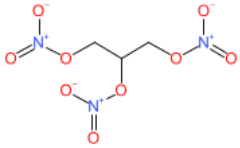



Prediction: Reliability:

Prediction is -1.48, 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)

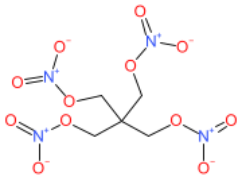
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Oral Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: -1.48
Predicted Oral Carcinogenicity SF [1/(mg/kg-day)]: 0.033
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



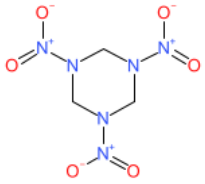
Compound #1

CAS: 55-63-0
Dataset id:218 (Training Set)
SMILES: O=[N+]([O-])OCC(O[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.609
Experimental value : -1.77
Predicted value : -1.69



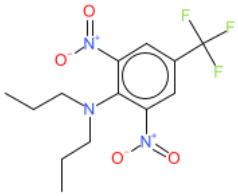
Compound #2

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: 0.601
Experimental value : -2.4
Predicted value : -2.004



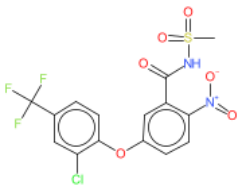
Compound #3

CAS: 121-82-4
Dataset id:173 (Test Set)
SMILES: O=[N+]([O-])N1CN([N+](=O)[O-])CN([N+](=O)[O-])C1
Similarity: 0.593
Experimental value : -0.96
Predicted value : -0.223



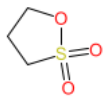
Compound #4

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: 0.577
Experimental value : -2.11
Predicted value : -1.41



Compound #5

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: 0.572
Experimental value : -0.72
Predicted value : -1.975



Compound #6

CAS: 1120-71-4
Dataset id:266 (Test Set)
SMILES: O=S1(=O)(OCCC1)
Similarity: 0.57
Experimental value : 0.38
Predicted value : -0.402

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.206<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ✖ | Similar molecules with known experimental value<br>Similarity index = 0.605<br>Explanation: No similar compounds with known experimental value in the training set have been found.. |
| ✔ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.238<br>Explanation: Accuracy of prediction for similar molecules found in the training set is good.. |
| ✔ | Concordance for similar molecules<br>Concordance index = 0.603<br>Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.. |
| ✔ | Maximum error of prediction among similar molecules<br>Max error index = 0.396<br>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<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.34<br>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.

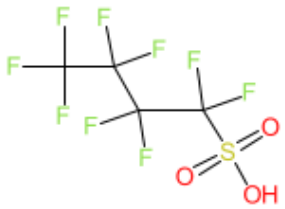
## 4.1 Reasoning:
### Relevant Chemical Fragments and Moieties

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

# 1. Prediction Summary

Prediction for compound Molecule 0 -

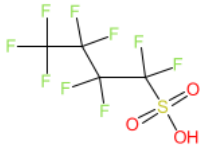



**EXPERIMENTAL DATA**

**E xperimental value is NON-Carcinogen. Model prediction is NON-Carcinogen (GOOD reliability).**

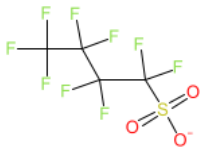
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: NON-Carcinogen
Predicted Inhalation Carcinogenic class: NON-Carcinogen
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



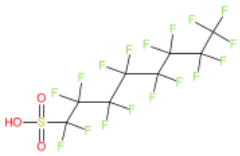
Compound #1

CAS: 375-73-5
Dataset id:613 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 1
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #2

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: 0.983
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

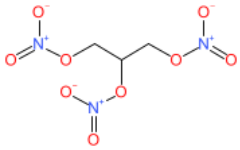


Compound #3

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: 0.809
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

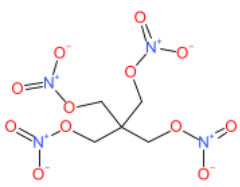
Compound #4

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: 0.732
Experimental value : NON-Carcinogen
Predicted value : Carcinogen



Compound #5

CAS: 55-63-0
Dataset id:587 (Training Set)
SMILES: O=[N+]([O-])OCC(O[N+](=O)[O-])CO[N+](=O)[O-]
Similarity: 0.609
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #6

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: 0.601
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✓ | Global AD Index<br>AD index = 0.85<br>Explanation: The predicted compound is into the Applicability Domain of the model. |
| ✓ | Similar molecules with known experimental value<br>Similarity index = 1<br>Explanation: Strongly similar compounds with known experimental value in the training set have been .. |
| ✓ | Accuracy of prediction for similar molecules<br>Accuracy index = 1<br>Explanation: Accuracy of prediction for similar molecules found in the training set is good.. |
| ✓ | Concordance for similar molecules<br>Concordance index = 1<br>Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.. |
| ✓ | Model's descriptors range check<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ⚠ | Atom Centered Fragments similarity check<br>ACF index = 0.85<br>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.

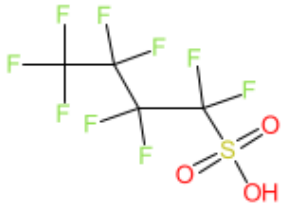
## 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

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

# 1. Prediction Summary

Prediction for compound Molecule 0 -

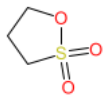



Prediction: Reliability:

Prediction is -0.42, 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 found)

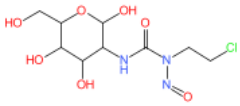
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted Inhalation Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: -0.42
Predicted Inhalation Carcinogenicity SF [1/(mg/kg-day)]: 0.3787
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



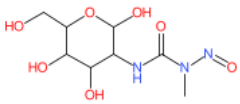
Compound #1

CAS: 1120-71-4
Dataset id:226 (Training Set)
SMILES: O=S1(=O)(OCCC1)
Similarity: 0.57
Experimental value : 0.38
Predicted value : -0.413



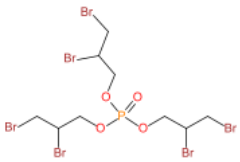
Compound #2

CAS: 54749-90-5
Dataset id:64 (Training Set)
SMILES: O=NN(C(=O)NC1C(O)OC(CO)C(O)C1(O))CCCl
Similarity: 0.568
Experimental value : 2.38
Predicted value : 1.702



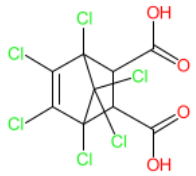
Compound #3

CAS: 18883-66-4
Dataset id:234 (Test Set)
SMILES: O=NN(C(=O)NC1C(O)OC(CO)C(O)C1(O))C
Similarity: 0.567
Experimental value : 2.04
Predicted value : 0.989



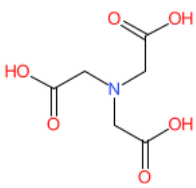
Compound #4

CAS: 126-72-7
Dataset id:257 (Training Set)
SMILES: O=P(OCC(CBr)Br)(OCC(CBr)Br)OCC(CBr)Br
Similarity: 0.563
Experimental value : 0.36
Predicted value : 0.569



Compound #5

CAS: 115-28-6
Dataset id:53 (Test Set)
SMILES: O=C(O)C1C(C(=O)O)C2(C(=C(C1(C2(Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.555
Experimental value : -1.04
Predicted value : 1.623



Compound #6

CAS: 139-13-9
Dataset id:178 (Training Set)
SMILES: O=C(O)CN(CC(=O)O)CC(=O)O
Similarity: 0.555
Experimental value : -2.28
Predicted value : 0.003

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✗ | Global AD Index<br>AD index = 0.228<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ✗ | Similar molecules with known experimental value<br>Similarity index = 0.569<br>Explanation: No similar compounds with known experimental value in the training set have been found.. |
| ✓ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.736<br>Explanation: Accuracy of prediction for similar molecules found in the training set is good.. |
| ✗ | Concordance for similar molecules<br>Concordance index = 1.802<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✓ | Maximum error of prediction among similar molecules<br>Max error index = 0.793<br>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<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ✗ | Atom Centered Fragments similarity check<br>ACF index = 0.4<br>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.

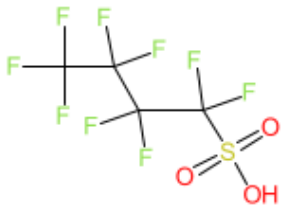
## 4.1 Reasoning:
### Relevant Chemical Fragments and Moieties

(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<br>The fragment has never been found in the model's training set |
|  | Fragment defined by the SMILES: CC(F)(F)S<br>The fragment has never been found in the model's training set |
|  | Fragment defined by the SMILES: CF<br>The fragment has never been found in the model's training set |
|  | Fragment defined by the SMILES: CC(F)(F)F<br>The fragment has never been found in the model's training set |

# 1. Prediction Summary

Prediction for compound Molecule 0 -

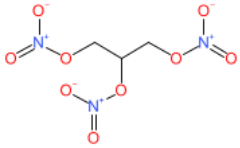



Prediction: Reliability:

Prediction is 1.5923, 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 high value, considering the experimental variability
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (4 unknown fragments and 2 infrequent_fragments found)

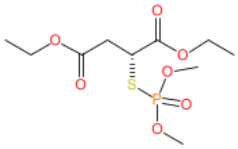
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: 1.5923
Predicted TD50 [mg/kg bw/day]: 0.0256
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



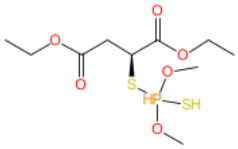
Compound #1

CAS: N.A.
Dataset id:24 (Training Set)
SMILES: C(C(CON(=O)=O)ON(=O)=O)ON(=O)=O
Similarity: 0.609
Experimental value : -2.344
Predicted value : -0.701



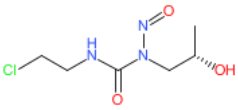
Compound #2

CAS: N.A.
Dataset id:38 (Training Set)
SMILES: CCOC(=O)[C@@H](CC(=O)OCC)SP(=O)(OC)OC
Similarity: 0.583
Experimental value : -3.193
Predicted value : -3.699



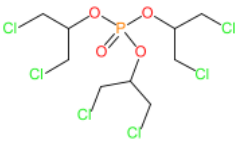
Compound #3

CAS: N.A.
Dataset id:26 (Training Set)
SMILES: P(S)(S[C@H](C(=O)OCC)CC(=O)OCC)(OC)OC
Similarity: 0.572
Experimental value : -2.614
Predicted value : -3.531



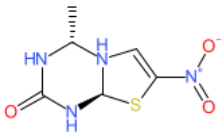
Compound #4

CAS: N.A.
Dataset id:155 (Test Set)
SMILES: N(C[C@@H](O)C)(C(=O)NCCCl)N=O
Similarity: 0.572
Experimental value : 0.065
Predicted value : -0.312



Compound #5

CAS: N.A.
Dataset id:18 (Training Set)
SMILES: P(=O)(OC(CCl)CCl)(OC(CCl)CCl)OC(CCl)CCl
Similarity: 0.566
Experimental value : -2.158
Predicted value : -2.535



Compound #6

CAS: N.A.
Dataset id:109 (Training Set)
SMILES: [C@@H]12SC(=C[NH]1[C@H](NC(=O)N2)C)N(=O)=O
Similarity: 0.566
Experimental value : -2.706
Predicted value : -2.609

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✗ | Global AD Index<br>AD index = 0.203<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ✗ | Similar molecules with known experimental value<br>Similarity index = 0.596<br>Explanation: No similar compounds with known experimental value in the training set have been found.. |
| ⚠ | Accuracy of prediction for similar molecules<br>Accuracy index = 1.075<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal.. |
| ✗ | Concordance for similar molecules<br>Concordance index = 4.361<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✗ | Maximum error of prediction among similar molecules<br>Max error index = 1.643<br>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<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ✗ | Atom Centered Fragments similarity check<br>ACF index = 0.34<br>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:

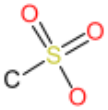
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

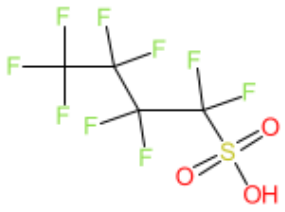
# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

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

# 1. Prediction Summary

Prediction for compound Molecule 0 -

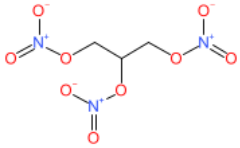



Prediction: Reliability:

Prediction is -5.0117, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (5 unknown fragments and 2 infrequent_fragments found)

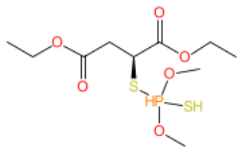
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: -5.0117
Predicted TD50 [mg/kg bw/day]: 102728.15
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



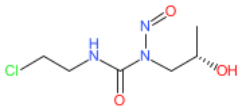
Compound #1

CAS: N.A.
Dataset id:94 (Training Set)
SMILES: C(C(CON(=O)=O)ON(=O)=O)ON(=O)=O
Similarity: 0.609
Experimental value : -2.517
Predicted value : -6.738



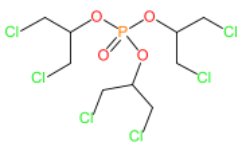
Compound #2

CAS: N.A.
Dataset id:107 (Training Set)
SMILES: P(S)(S[C@H](C(=O)OCC)CC(=O)OCC)(OC)OC
Similarity: 0.572
Experimental value : -3.588
Predicted value : -2.078



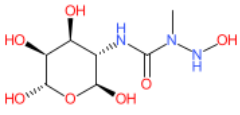
Compound #3

CAS: N.A.
Dataset id:152 (Test Set)
SMILES: N(C[C@@H](O)C)(C(=O)NCCCl)N=O
Similarity: 0.572
Experimental value : 0.053
Predicted value : -0.968



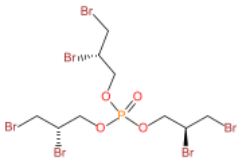
Compound #4

CAS: N.A.
Dataset id:92 (Training Set)
SMILES: P(=O)(OC(CCl)CCl)(OC(CCl)CCl)OC(CCl)CCl
Similarity: 0.566
Experimental value : -2.314
Predicted value : -1.883



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)NC(=O)N(NO)C)O
Similarity: 0.565
Experimental value : -0.646
Predicted value : 3.177



Compound #6

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: 0.563
Experimental value : -3.45
Predicted value : -3.074

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.2<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ✖ | Similar molecules with known experimental value<br>Similarity index = 0.59<br>Explanation: No similar compounds with known experimental value in the training set have been found.. |
| ✖ | Accuracy of prediction for similar molecules<br>Accuracy index = 2.865<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 1.959<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Maximum error of prediction among similar molecules<br>Max error index = 4.221<br>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<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ✖ | Atom Centered Fragments similarity check<br>ACF index = 0.34<br>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:

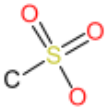
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

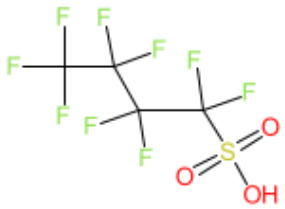
# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

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

# 1. Prediction Summary

Prediction for compound Molecule 0 -

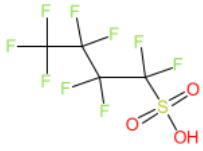



**EXPERIMENTAL DATA**

**E xperimental value is 433.84 mg/Kg. Model prediction is 433.84 mg/kg (LOW reliability).**

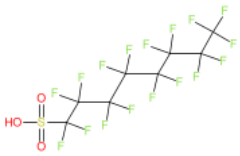
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: 0.16
Predicted log LD50 [log(mmol/Kg)]: 0.16
Predicted log LD50 [mg/Kg]: 433.84
Molecules used for prediction: 1
Experimental value [mg/Kg]: 433.84
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



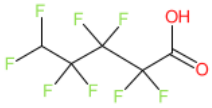
Compound #1

CAS: N.A.
Dataset id:3431 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 1
Experimental value : 0.16
Predicted value : -



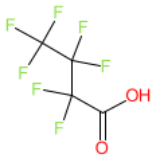
Compound #2

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)C(F)(F)F
Similarity: 0.809
Experimental value : -0.52
Predicted value : -0.203



Compound #3

CAS: N.A.
Dataset id:3530 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)F
Similarity: 0.787
Experimental value : 0.85
Predicted value : 0.786

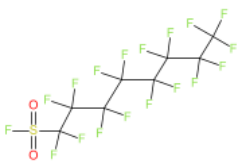


Compound #4

CAS: N.A.
Dataset id:3430 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.785
Experimental value : 0.82
Predicted value : 0.808

Compound #5

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: 0.764
Experimental value : 0.62
Predicted value : 0.397



Compound #6

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)C(F)(F)F
Similarity: 0.755
Experimental value : -0.51
Predicted value : -0.478

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.7<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ✔ | Similar molecules with known experimental value<br>Similarity index = 1<br>Explanation: Strongly similar compounds with known experimental value in the training set have been .. |
| ✖ | Accuracy of prediction for similar molecules<br>Accuracy index = 2.4<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate because all predictions are missing values.. |
| ✔ | Concordance for similar molecules<br>Concordance index = 0<br>Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.. |
| ✖ | Maximum error of prediction among similar molecules<br>Max error index = 2.4<br>Explanation: the maximum error in prediction can not be evaluated because all predictions are missing values.. |
| ✔ | Atom Centered Fragments similarity check<br>ACF index = 1<br>Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.. |

Symbols explanation:

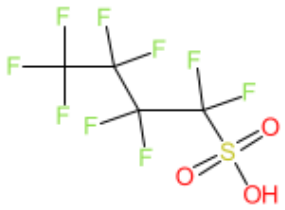
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 0.72 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
- 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)
The following relevant fragments have been found: 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)F
Experimental value: -
Predicted BCF [log(L/kg)]: 0.72
Predicted BCF [L/kg]: 5
Predicted BCF from sub-model 1 (HM) [log(L/kg)]: 0.88
Predicted BCF from sub-model 2 (GA) [log(L/kg)]: 0.93
Predicted LogP (MLogP): 1.68
Structural Alerts: SO3H group (PG 02)
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

## 2. Possible Use and Uncertainty

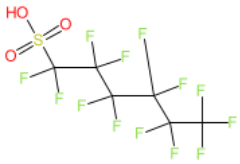
Threshold 3.3 (bioaccumulative)

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

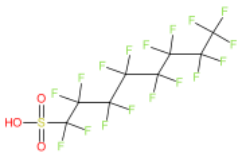


Compound #1

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: 0.885
Experimental value : 3.6
Predicted value : 1.585

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

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



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: 0.809
Experimental value : 3.73
Predicted value : 1.697

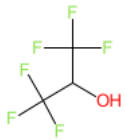
Alerts (found also in the target): SO3H group (PG 02)

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

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: 0.732
Experimental value : 3.12
Predicted value : 2.534

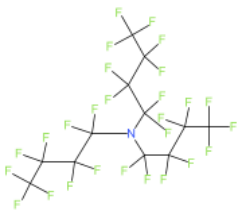
Alerts (not found also in the target): 10 F atoms in the molecule (SO 10); Carbonyl residue (SR 02); COOH group (PG 01)



Compound #4

CAS: 920-66-1
Dataset id:263 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.717
Experimental value : 0.3
Predicted value : 0.601

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

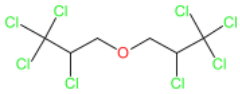


Compound #5

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: 0.596
Experimental value : 1.42
Predicted value : 0.879

Alerts (not found also in the target): 10 F atoms in the molecule (SO 10); Tertiary amine (SR 05)

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



Compound #6

CAS: 127-90-2
Dataset id:460 (Training Set)
SMILES: O(CC(C(Cl)(Cl)Cl)Cl)CC(C(Cl)(Cl)Cl)Cl
Similarity: 0.587
Experimental value : 3.28
Predicted value : 2.936

Alerts (not found also in the target): 6 Cl atoms in the molecule (SO 01)

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.716<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.842<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ✖ | Accuracy of prediction for similar molecules<br>Accuracy index = 2.024<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 2.942<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Maximum error of prediction among similar molecules<br>Max error index = 2.033<br>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<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.. |
| ⚠ | Atom Centered Fragments similarity check<br>ACF index = 0.85<br>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.

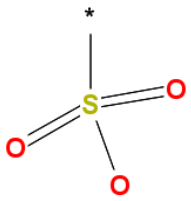
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

# 4.1 Reasoning:
## Relevant Chemical Fragments and Moieties

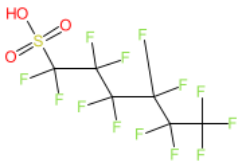
(Molecule 0) Reasoning on fragments/structural alerts :.

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.

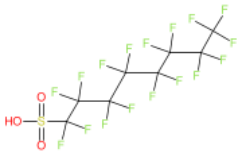


CAS: 355-46-4
Dataset id:55 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.885

Experimental value : 3.6
Predicted value : 1.585

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

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

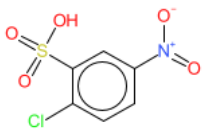


CAS: 1763-23-1
Dataset id:57 (Training Set)
SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.809

Experimental value : 3.73
Predicted value : 1.697

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

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



CAS: 96-73-1
Dataset id:474 (Training Set)
SMILES: O=[N+]([O-])c1ccc(c(c1)S(=O)(=O)O)Cl
Similarity: 0.576

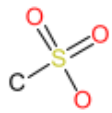
Experimental value : 0.48
Predicted value : 0.147

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

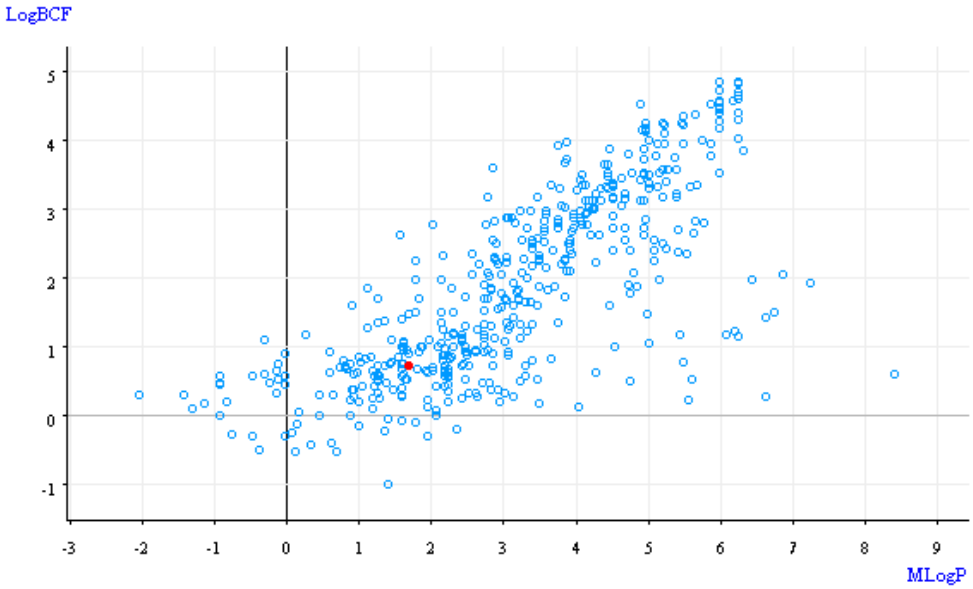
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

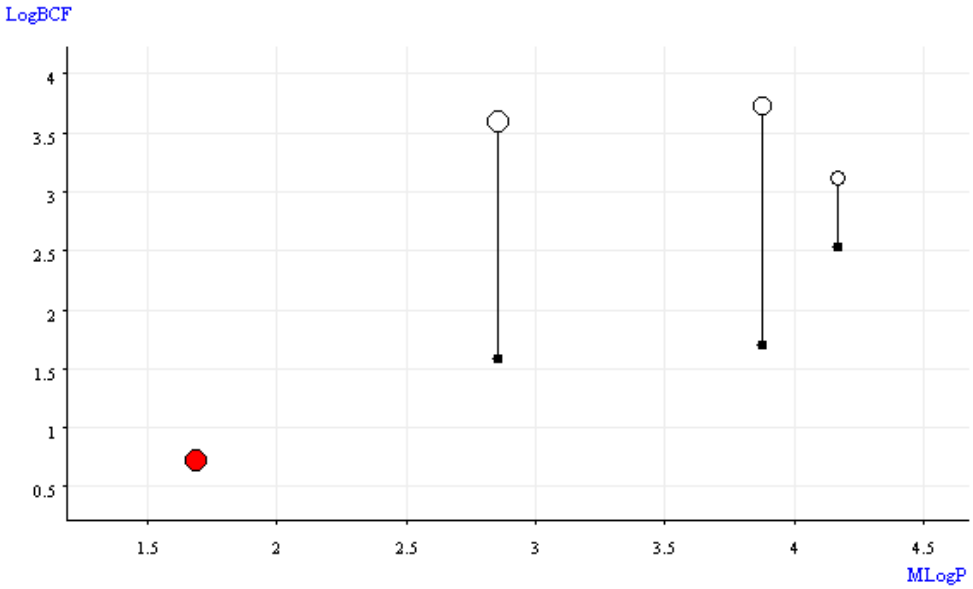
# 4.2 Reasoning:
## Analysis of Molecular Descriptors

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.

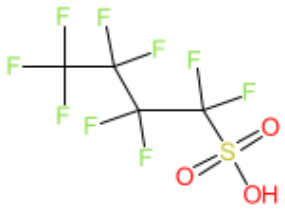



Following, a scatterplot of MLogP against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound;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



# 1. Prediction Summary

Prediction for compound Molecule 0 -

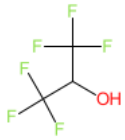



Prediction: Reliability:

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)

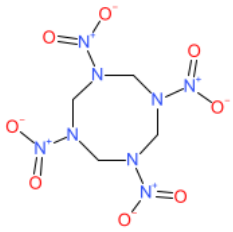
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF [log(L/kg)]: 0.5
Predicted BCF [L/kg]: 3
Predicted LogP (Meylan/Kowwin): 2.41
Predicted LogP reliability: Low
MW: 299.79
Ionic compound: yes
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



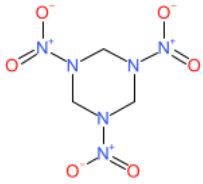
Compound #1

CAS: 920-66-1
Dataset id:117 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.717
Experimental value : 0.4
Predicted value : 0.762



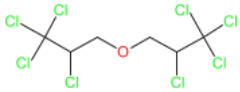
Compound #2

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: 0.593
Experimental value : -0.3
Predicted value : 2.139



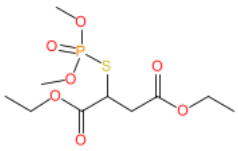
Compound #3

CAS: 121-82-4
Dataset id:90 (Training Set)
SMILES: O=[N+]([O-])N1CN([N+](=O)[O-])CN([N+](=O)[O-])C1
Similarity: 0.593
Experimental value : 0.3
Predicted value : 0.5



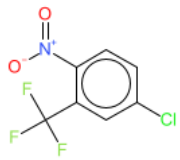
Compound #4

CAS: 127-90-2
Dataset id:406 (Training Set)
SMILES: O(CC(C(Cl)(Cl)Cl)Cl)CC(C(Cl)(Cl)Cl)Cl
Similarity: 0.587
Experimental value : 3.28
Predicted value : 3.034



Compound #5

CAS: 1634-78-2
Dataset id:82 (Training Set)
SMILES: O=C(OCC)CC(C(=O)OCC)SP(=O)(OC)OC
Similarity: 0.583
Experimental value : 0.05
Predicted value : 0.5



Compound #6

CAS: 118-83-2
Dataset id:240 (Training Set)
SMILES: O=[N+]([O-])c1ccc(cc1C(F)(F)F)Cl
Similarity: 0.581
Experimental value : 1.87
Predicted value : 1.778

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✗ | Global AD Index<br>AD index = 0.219<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ✗ | Similar molecules with known experimental value<br>Similarity index = 0.645<br>Explanation: No similar compounds with known experimental value in the training set have been found.. |
| ✗ | Accuracy of prediction for similar molecules<br>Accuracy index = 1.401<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate.. |
| ✓ | Concordance for similar molecules<br>Concordance index = 0.45<br>Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value.. |
| ✗ | Maximum error of prediction among similar molecules<br>Max error index = 2.439<br>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<br>LogP reliability = 0<br>Explanation: reliability of logP value used by the model is not adequate.. |
| ✓ | Model's descriptors range check<br>Descriptors range check = True<br>Explanation: descriptors for this compound have values inside the defined range.. |
| ✗ | Atom Centered Fragments similarity check<br>ACF index = 0.34<br>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:

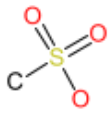
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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

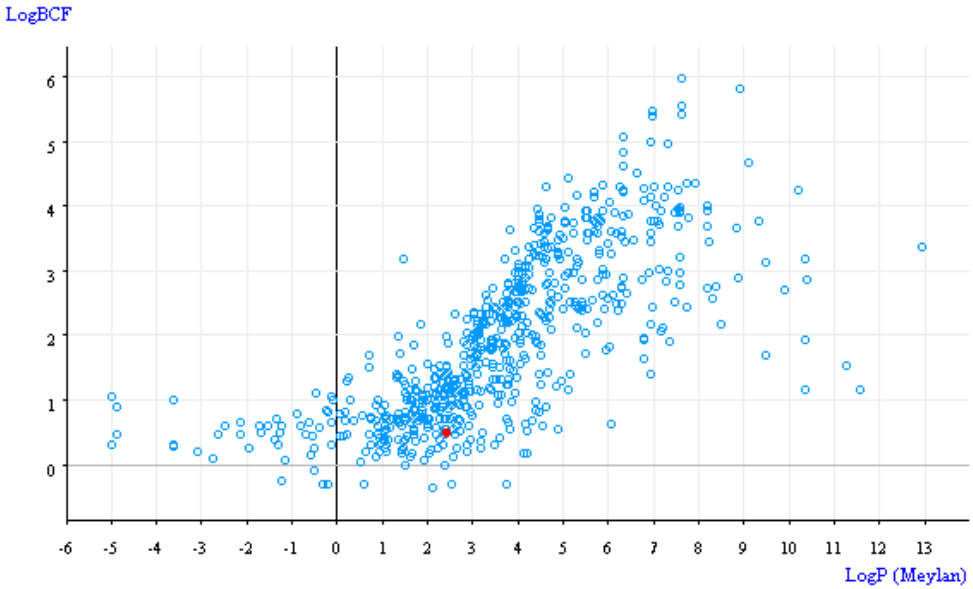
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

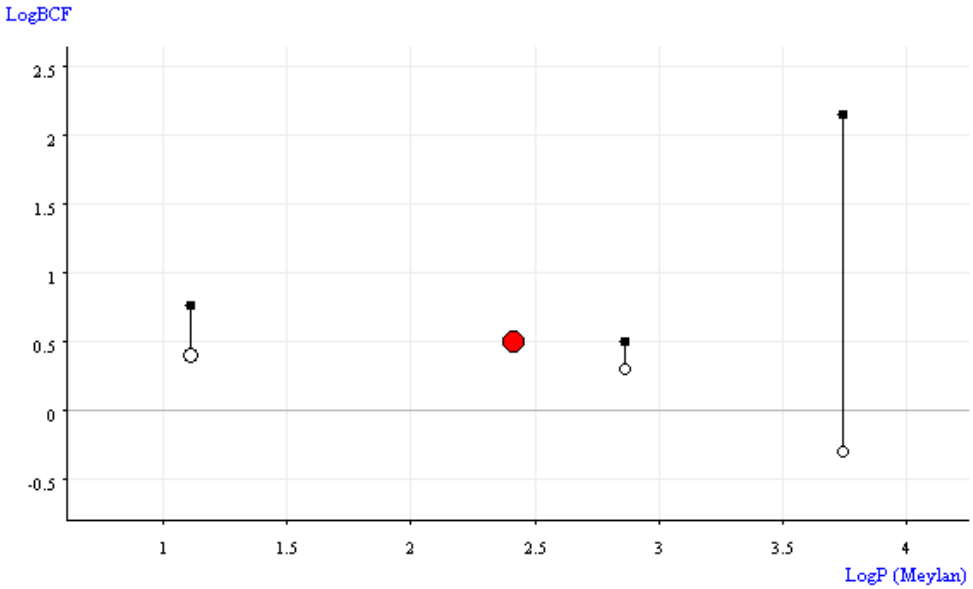
## 4.2 Reasoning:
### Analysis of Molecular Descriptors

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.

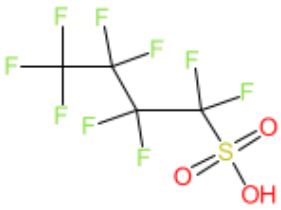



Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same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



# 1. Prediction Summary

Prediction for compound Molecule 0 -

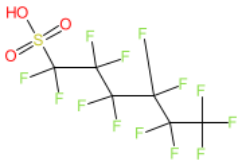



Prediction: Reliability:

Prediction is 1.4 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not 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 high value, considering the experimental variability
- reliability of logP value used by the model is not adequate
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent_fragments found)

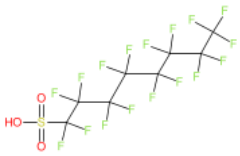
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF (up) [log(L/kg)]: 1.4
Predicted BCF (up) [L/kg]: 25
Predicted BCF (low) [log(L/kg)]: 1.19
Predicted BCF (low) [L/kg]: 16
Predicted BCF (mid) [log(L/kg)]: 1.24
Predicted BCF (mid) [L/kg]: 17
Predicted LogP (Meylan/Kowwin): 2.41
Predicted LogP reliability: Low
Predicted kM (Meylan): -0.23
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: 0.885
Experimental value : 1.62
Predicted value : 2.989

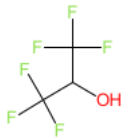


Compound #2

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: 0.809
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: 0.732
Experimental value : 0.977
Predicted value : 4



Compound #4

CAS: 920-66-1
Dataset id:191 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.717
Experimental value : 0.2
Predicted value : 0.565

Compound #5

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: 0.683
Experimental value : 3.04
Predicted value : 3.236

Compound #6

CAS: 2058-94-8
Dataset id:665 (Training Set)
SMILES: O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.662
Experimental value : 3.72
Predicted value : 2.565

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.716<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ⚠ | Similar molecules with known experimental value<br>Similarity index = 0.842<br>Explanation: Only moderately similar compounds with known experimental value in the training set have been found.. |
| ⚠ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.846<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal.. |
| ✖ | Concordance for similar molecules<br>Concordance index = 1.148<br>Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Maximum error of prediction among similar molecules<br>Max error index = 1.369<br>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<br>LogP reliability = 0<br>Explanation: reliability of logP value used by the model is not adequate.. |
| ⚠ | Atom Centered Fragments similarity check<br>ACF index = 0.85<br>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

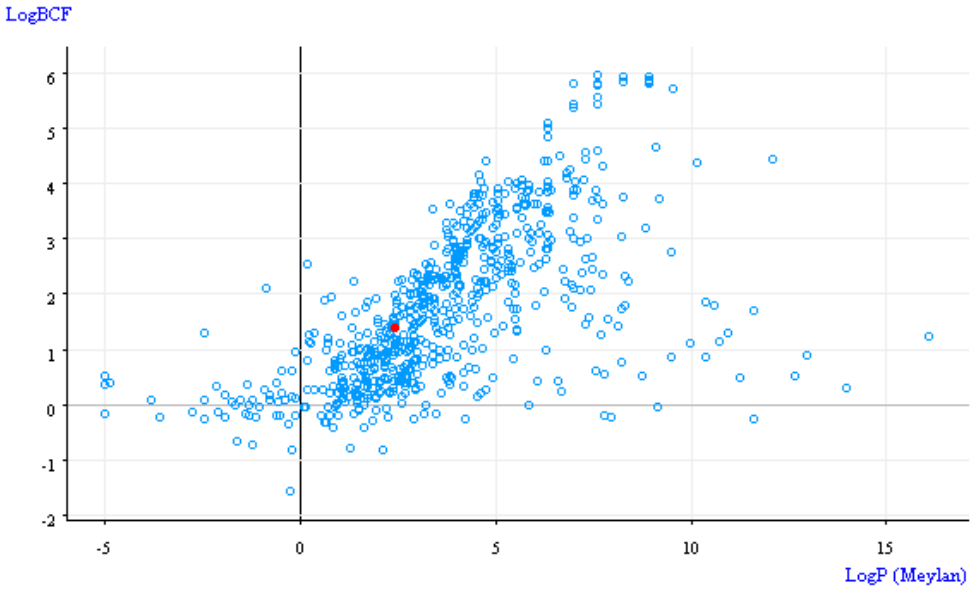
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:


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

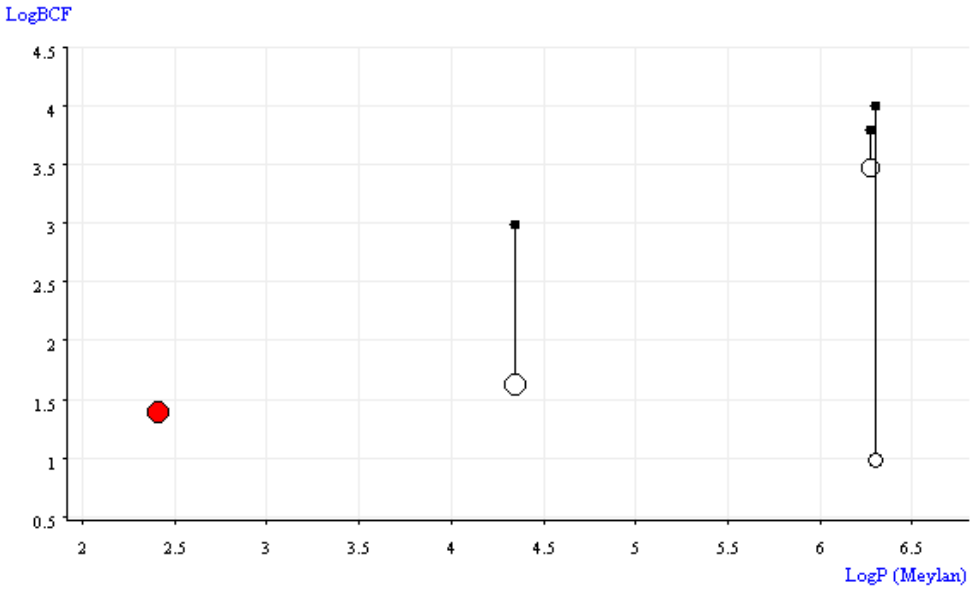
# 4.2 Reasoning:
## Analysis of Molecular Descriptors

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.

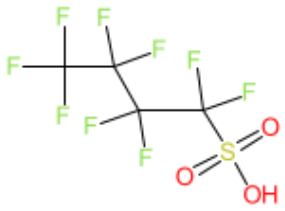



Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same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



# 1. Prediction Summary

Prediction for compound Molecule 0 -

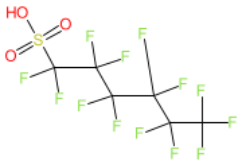



Prediction: Reliability:

Prediction is 2.92 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:
- 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
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
- 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)

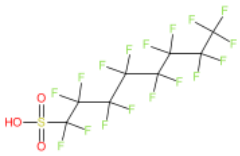
Compound: Molecule 0
Compound SMILES: O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Experimental value: -
Predicted BCF [log(L/kg)]: 2.92
Molecules used for prediction: 4
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



Compound #1

CAS: 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: 0.885
Experimental value : 3.6
Predicted value : 3.479

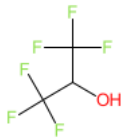


Compound #2

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: 0.809
Experimental value : 3.73
Predicted value : 2.948

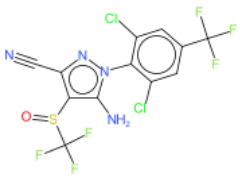
Compound #3

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: 0.732
Experimental value : 3.12
Predicted value : 3.096



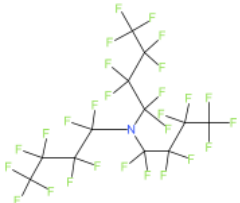
Compound #4

CAS: 920-66-1
Dataset id:391 (Training Set)
SMILES: FC(F)(F)C(O)C(F)(F)F
Similarity: 0.717
Experimental value : 0.244
Predicted value : 1.753



Compound #5

CAS: 120068-37-3
Dataset id:858 (Training Set)
SMILES: N#Cc1nn(c(N)c1S(=O)C(F)(F)F)c2c(cc(cc2Cl)C(F)(F)F)Cl
Similarity: 0.601
Experimental value : 2.507
Predicted value : 1.969



Compound #6

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)C(F)(F)F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
Similarity: 0.596
Experimental value : 1.299
Predicted value : 3.432

## 3.2 Applicability Domain:
### Measured Applicability Domain Scores

| | |
|---|---|
| ✖ | Global AD Index<br>AD index = 0.649<br>Explanation: The predicted compound is outside the Applicability Domain of the model. |
| ✔ | Similar molecules with known experimental value<br>Similarity index = 0.764<br>Explanation: Strongly similar compounds with known experimental value in the training set have been .. |
| ⚠ | Accuracy of prediction for similar molecules<br>Accuracy index = 0.609<br>Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal.. |
| ⚠ | Concordance for similar molecules<br>Concordance index = 1.093<br>Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value.. |
| ✖ | Maximum error of prediction among similar molecules<br>Max error index = 1.509<br>Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability.. |
| ⚠ | Atom Centered Fragments similarity check<br>ACF index = 0.85<br>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