VEGA QSAR MODELS

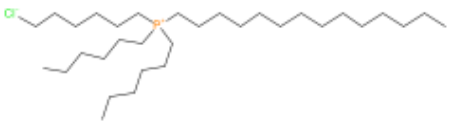
Report

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 -



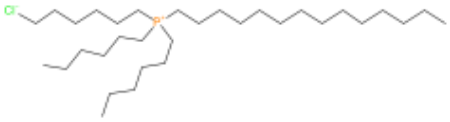
Prediction:

Prediction is Mutagenic with a consensus score of 0.15, based on 4 models.

Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Used models: 4
Predicted Consensus Mutagen activity: Mutagenic
Mutagenic Score: 0.15
Non-Mutagenic Score: 0.05
Model Caesar assessment: Suspect Mutagenic (LOW reliability)
Model ISS assessment: Mutagenic (LOW reliability)
Model SarPy assessment: Mutagenic (LOW reliability)
Model KNN assessment: NON-Mutagenic (LOW reliability)
Remarks:
none

1. Prediction Summary

Prediction for compound Molecule 0 -



Prediction: Reliability:

Prediction is Suspect Mutagenic, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- 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
- 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: SA8 Aliphatic halogens

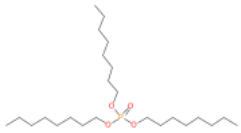
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Mutagen activity: Suspect Mutagenic
Structural Alerts: SA8 Aliphatic halogens
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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

Compound #1

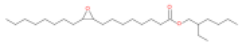
CAS: 112-52-7
Dataset id:3322 (Test Set)
SMILES: CCCCCCCCCCCCCl
Similarity: 0.691
Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

Alerts (found also in the target): SA8 Aliphatic halogens



Compound #2

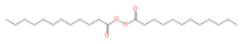
CAS: 1806-54-8
Dataset id:3666 (Training Set)
SMILES: O=P(OCCCCCCCC)(OCCCCCCCC)OCCCCCCCC
Similarity: 0.682
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #3

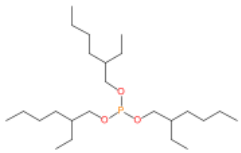
CAS: 141-38-8
Dataset id:2569 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCCCCC1OC1(CCCCCCCC)
Similarity: 0.674
Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

Alerts (not found also in the target): SA7 Epoxides and aziridines



Compound #4

CAS: 105-74-8
Dataset id:1470 (Training Set)
SMILES: O=C(OOC(=O)CCCCCCCCCCC)CCCCCCCCCCC
Similarity: 0.673
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #5

CAS: 301-13-3
Dataset id:3329 (Training Set)
SMILES: O(CC(CC)CCCC)P(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.668
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #6

CAS: 110-30-5
Dataset id:2884 (Training Set)
SMILES: O=C(NCCNC(=O)CCCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCCCC
Similarity: 0.666
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.682 Explanation: No similar compounds with known experimental value in the training set have been found..
✗	Accuracy of prediction for similar molecules Accuracy index = 0.333 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✗	Concordance for similar molecules Concordance index = 0 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
⚠	Atom Centered Fragments similarity check ACF index = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)..

Symbols explanation:

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

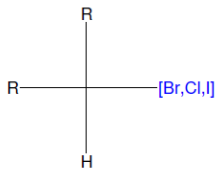
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

Fragment found: SA8 Aliphatic halogens



R = any atom/group

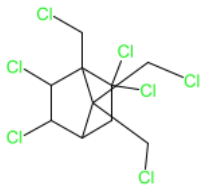
This alert contains non tertiary aliphatic halogens. Substances fired by Alerts SA2, SA4, SA5 and SA20 should be also excluded.

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

CAS: 112-52-7
Dataset id:3322 (Test Set)
SMILES: CCCCCCCCCCCCCl
Similarity: 0.691

Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

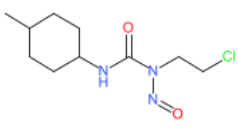
Alerts (found also in the target): SA8 Aliphatic halogens



CAS: 51775-36-1
Dataset id:3431 (Training Set)
SMILES: C2C1C(C(C(CCl)(C1(CCl)(CCl))C2(Cl)Cl)Cl)Cl
Similarity: 0.527

Experimental value : NON-Mutagenic
Predicted value : Suspect Mutagenic

Alerts (found also in the target): SA8 Aliphatic halogens



CAS: 13909-09-6
Dataset id:2393 (Test Set)
SMILES: O=NN(C(=O)NC1CCC(C)CC1)CCCl
Similarity: 0.517

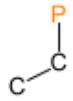
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SA8 Aliphatic halogens

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

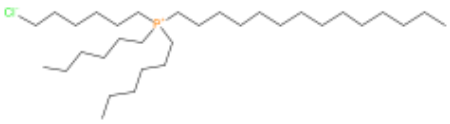
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

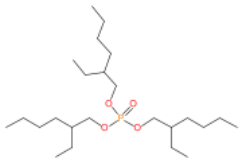


Prediction: Reliability:

Prediction is 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
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)
The following alerts have been found: SA8 Aliphatic halogens

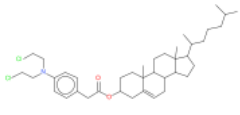
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Mutagen activity: Mutagenic
Structural Alerts: SA8 Aliphatic halogens
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



Compound #1

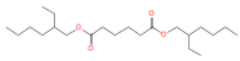
CAS: 78-42-2
Dataset id:69 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #2

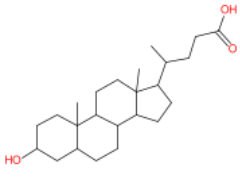
CAS: 3546-10-9
Dataset id:216 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA5 S or N mustard



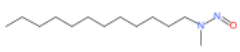
Compound #3

CAS: 103-23-1
Dataset id:52 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

CAS: 434-13-9
Dataset id:117 (Training Set)
SMILES: O=C(O)CCC(C)C2CCC3C4CCC1CC(O)CCC1(C)C4(CCC23(C))
Similarity: 0.607
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

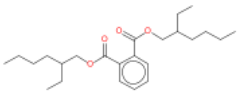


Compound #5

CAS: 55090-44-3
Dataset id:547 (Training Set)
SMILES: O=NN(C)CCCCCCCCCCCC
Similarity: 0.571
Experimental value : Mutagenic
Predicted value : Mutagenic

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: 117-81-7
Dataset id:53 (Training Set)
SMILES: O=C(OCC(CC)CCCC)c1ccccc1(C(=O)OCC(CC)CCCC)
Similarity: 0.566
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.637 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.521 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✖	Concordance for similar molecules Concordance index = 0 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.51 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)..

Symbols explanation:

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

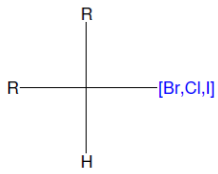
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

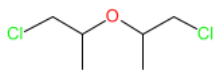
Fragment found: SA8 Aliphatic halogens



R = any atom/group

This alert contains non tertiary aliphatic halogens. Substances fired by Alerts SA2, SA4, SA5 and SA20 should be also excluded.

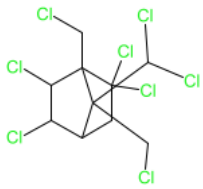
Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 108-60-1
Dataset id:40 (Training Set)
SMILES: O(C(C)CCl)C(C)CCl
Similarity: 0.526

Experimental value : Mutagenic
Predicted value : Mutagenic

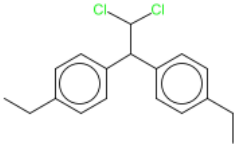
Alerts (found also in the target): SA8 Aliphatic halogens



CAS: 8001-35-2
Dataset id:727 (Training Set)
SMILES: C2C1C(C(C(CCl)(C1(CCl)C(Cl)Cl)C2(Cl)Cl)Cl)Cl
Similarity: 0.522

Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SA8 Aliphatic halogens



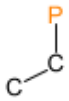
CAS: 72-56-0
Dataset id:192 (Training Set)
SMILES: c1cc(ccc1CC)C(c2ccc(cc2)CC)C(Cl)Cl
Similarity: 0.515

Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SA8 Aliphatic halogens

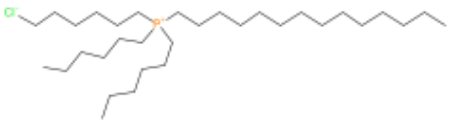
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(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: CP(C)(C)C The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: CCP 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 adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)
The following relevant fragments have been found: SM106; SM157; SM163; SM185

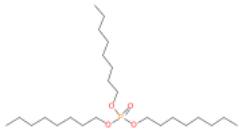
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Mutagen activity: Mutagenic
No. alerts for mutagenicity: 1
No. alerts for non-mutagenicity: 3
Structural Alerts: SM106; SM157; SM163; SM185
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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

Compound #1

CAS: 112-52-7
Dataset id:3322 (Test Set)
SMILES: CCCCCCCCCCCCCl
Similarity: 0.691
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM106; SM157; SM163

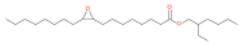


Compound #2

CAS: 1806-54-8
Dataset id:3666 (Training Set)
SMILES: O=P(OCCCCCCCC)(OCCCCCCCC)OCCCCCCCC
Similarity: 0.682
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM163

Alerts (not found also in the target): SM124

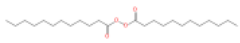


Compound #3

CAS: 141-38-8
Dataset id:2569 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCCCCC1OC1(CCCCCCCC)
Similarity: 0.674
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM157; SM163

Alerts (not found also in the target): SM97; SM123; SM143; SM169; SM177; SM178; SM182; SM188

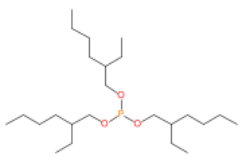


Compound #4

CAS: 105-74-8
Dataset id:1470 (Training Set)
SMILES: O=C(OOC(=O)CCCCCCCCCCC)CCCCCCCCCCC
Similarity: 0.673
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM157; SM163

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



Compound #5

CAS: 301-13-3
Dataset id:3329 (Training Set)
SMILES: O(CC(CC)CCCC)P(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.668
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM163

Alerts (not found also in the target): SM124; SM169; SM182

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

Compound #6

CAS: 110-30-5
Dataset id:2884 (Training Set)
SMILES: O=C(NCCNC(=O)CCCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCCCC
Similarity: 0.666
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (found also in the target): SM157; SM163

Alerts (not found also in the target): SM177

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.682 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 0.333 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✖	Concordance for similar molecules Concordance index = 0 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Atom Centered Fragments similarity check ACF index = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)..

Symbols explanation:

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

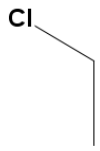
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on fragments/structural alerts - 1 of 4:.

Fragment found: SM106



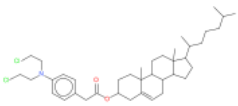
Sarpy alert n. 106 for Mutagenicity, defined by SMARTS: CCCl

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

CAS: 112-52-7
Dataset id:3322 (Test Set)
SMILES: CCCCCCCCCCCCCl
Similarity: 0.691

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM106; SM157; SM163

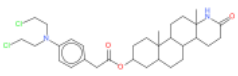


CAS: 3546-10-9
Dataset id:4248 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM106; SM157; SM163

Alerts (not found also in the target): SM45; SM73; SM104; SM162; SM169; SM182; SM188; SM195



CAS: 43000-65-3
Dataset id:4260 (Test Set)
SMILES:
O=C(OC4CCC2(C)(C(CCC1C3CCC(=O)NC3(C)(CCC12))C4))Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.596

Experimental value : Mutagenic
Predicted value : Mutagenic

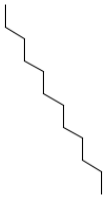
Alerts (found also in the target): SM106; SM157; SM163

Alerts (not found also in the target): SM45; SM73; SM104; SM162; SM169; SM177; SM182; SM188; SM195

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on fragments/structural alerts - 2 of 4:.

Fragment found: SM157



Sarpy alert n. 157 for NON-Mutagenicity, defined by SMARTS: CCCCCCCCCCCC

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

CAS: 112-52-7
Dataset id:3322 (Test Set)
SMILES: CCCCCCCCCCCCCl
Similarity: 0.691

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

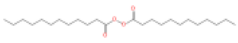
Alerts (found also in the target): SM106; SM157; SM163

CAS: 141-38-8
Dataset id:2569 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCCCCC1OC1(CCCCCCCC)
Similarity: 0.674

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM157; SM163

Alerts (not found also in the target): SM97; SM123; SM143; SM169; SM177; SM178; SM182; SM188



CAS: 105-74-8
Dataset id:1470 (Training Set)
SMILES: O=C(OOC(=O)CCCCCCCCCCC)CCCCCCCCCCC
Similarity: 0.673

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

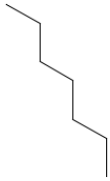
Alerts (found also in the target): SM157; SM163

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on fragments/structural alerts - 3 of 4:.

Fragment found: SM163



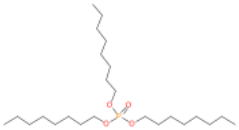
Sarpy alert n. 163 for NON-Mutagenicity, defined by SMARTS: CCCCCCC

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

CAS: 112-52-7
Dataset id:3322 (Test Set)
SMILES: CCCCCCCCCCCCCl
Similarity: 0.691

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM106; SM157; SM163



CAS: 1806-54-8
Dataset id:3666 (Training Set)
SMILES: O=P(OCCCCCCCC)(OCCCCCCCC)OCCCCCCCC
Similarity: 0.682

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

Alerts (found also in the target): SM163

Alerts (not found also in the target): SM124

CAS: 141-38-8
Dataset id:2569 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCCCCC1OC1(CCCCCCCC)
Similarity: 0.674

Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM157; SM163

Alerts (not found also in the target): SM97; SM123; SM143; SM169; SM177; SM178; SM182; SM188

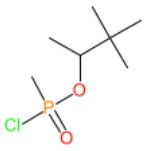
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on fragments/structural alerts - 4 of 4:.

Fragment found: SM185

Sarpy alert n. 185 for NON-Mutagenicity, defined by SMARTS: PC

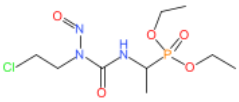
Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 7040-57-5
Dataset id:3926 (Training Set)
SMILES: O=P(OC(C)C(C)(C)C)(C)Cl
Similarity: 0.51

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

Alerts (found also in the target): SM185

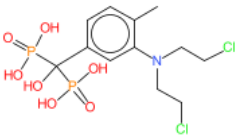


CAS: 92118-27-9
Dataset id:2445 (Training Set)
SMILES: O=NN(C(=O)NC(C)P(=O)(OCC)OCC)CCCl
Similarity: 0.479

Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM106; SM185

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



CAS: 106871-13-0
Dataset id:972 (Training Set)
SMILES: O=P(O)(O)C(O)(c1ccc(c(c1)N(CCCl)CCCl)C)P(=O)(O)O
Similarity: 0.444

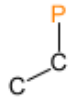
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (found also in the target): SM106; SM185

Alerts (not found also in the target): SM45; SM73

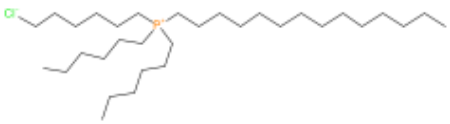
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(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: CCP The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: CP(C)(C)C 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 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 and 1 infrequent_fragments found)

Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Molecules used for prediction: 2
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

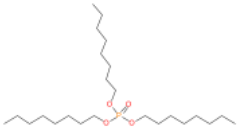
CAS: 2598-99-4
Dataset id:2466 (Training Set)
SMILES: O=C(OCCCCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCC
Similarity: 0.719
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #2

CAS: 557-61-9
Dataset id:3643 (Training Set)
SMILES: OCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Similarity: 0.711
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #3

CAS: 112-52-7
Dataset id:603 (Training Set)
SMILES: CCCCCCCCCCCCCl
Similarity: 0.691
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

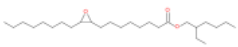


Compound #4

CAS: 1806-54-8
Dataset id:1965 (Training Set)
SMILES: O=P(OCCCCCCCC)(OCCCCCCCC)OCCCCCCCC
Similarity: 0.682
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #5

CAS: 10094-45-8
Dataset id:50 (Training Set)
SMILES: O=C(NCCCCCCCCCCCCCCCCC)CCCCCCCCCCCC=CCCCCCCCC
Similarity: 0.676
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #6

CAS: 141-38-8
Dataset id:1365 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCCCCC1OC1(CCCCCCCC)
Similarity: 0.674
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.431 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.715 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✗	Atom Centered Fragments similarity check ACF index = 0.51 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)..

Symbols explanation:

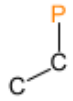
✓ The feature has a good assessment, model is reliable regarding this aspect.

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

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

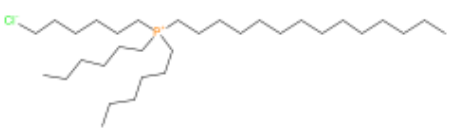
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

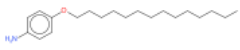


Prediction: Reliability:

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

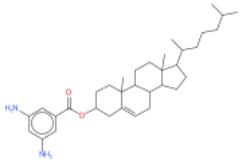
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



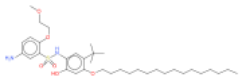
Compound #1

CAS: N.A.
Dataset id:6537 (Training Set)
SMILES: O(c1ccc(N)cc1)CCCCCCCCCCCCCC
Similarity: 0.599
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



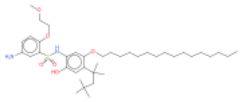
Compound #2

CAS: N.A.
Dataset id:7188 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)c5cc(N)cc(N)c5
Similarity: 0.58
Experimental value : NON-Mutagenic
Predicted value : NA



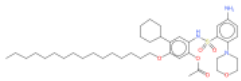
Compound #3

CAS: N.A.
Dataset id:4037 (Training Set)
SMILES:
O=S(=O)(Nc1cc(c(OCCCCCCCCCCCCCCCC)cc1(O))C(C)(C)C)c2cc(N)ccc2(OCCOC)
Similarity: 0.56
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

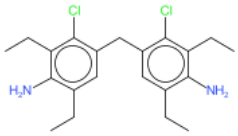
CAS: N.A.
Dataset id:6841 (Training Set)
SMILES:
O=S(=O)(Nc1cc(OCCCCCCCCCCCCCCCC)c(cc1(O))C(C)(C)CC(C)(C)C)c2cc(N)ccc2(OCCOC)
Similarity: 0.554
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #5

CAS: N.A.
Dataset id:4359 (Training Set)
SMILES:
O=C(Oc3cc(OCCCCCCCCCCCCCCCC)c(cc3(NS(=O)(=O)c1cc(N)ccc1N2CCOCC2))C4CCCCC4)C
Similarity: 0.544
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

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



Compound #6

CAS: N.A.
Dataset id:6633 (Training Set)
SMILES: Nc1c(cc(c(c1CC)Cl)Cc2cc(c(N)c(c2Cl)CC)CC)CC
Similarity: 0.54
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.589 Explanation: No similar compounds with known experimental value in the training set have been found..
✗	Accuracy of prediction for similar molecules Accuracy index = 0 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✓	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✗	Atom Centered Fragments similarity check ACF index = 0.34 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments and 1 infrequent_fragments found)..

Symbols explanation:

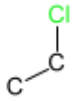
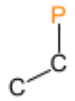
✓ The feature has a good assessment, model is reliable regarding this aspect.

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

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

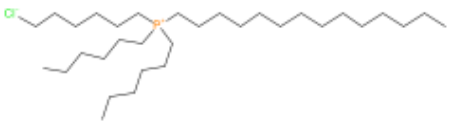
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -

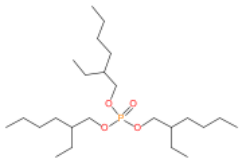


Prediction: Reliability:

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
- 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)

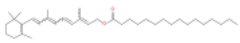
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Carcinogen activity: NON-Carcinogen
P(Carcinogen): 0.366
P(NON-Carcinogen): 0.634
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



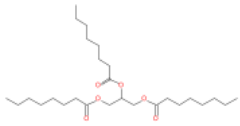
Compound #1

CAS: 78-42-2
Dataset id:784 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



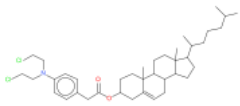
Compound #2

CAS: 79-81-2
Dataset id:693 (Training Set)
SMILES: O=C(OCC=C(C=CC=C(C=CC1=C(C)CCCC1(C)C)C)C)CCCCCCCCCCCCCCCC
Similarity: 0.649
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



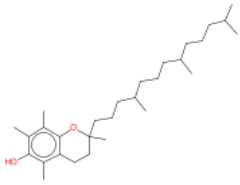
Compound #3

CAS: 538-23-8
Dataset id:759 (Test Set)
SMILES: O=C(OCC(OC(=O)CCCCCCC)COC(=O)CCCCCCC)CCCCCCC
Similarity: 0.631
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



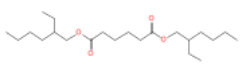
Compound #4

CAS: 3546-10-9
Dataset id:636 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : Carcinogen
Predicted value : Carcinogen



Compound #5

CAS: 10191-41-0
Dataset id:749 (Training Set)
SMILES: Oc2c(c(c1OC(C)(CCc1c2C)CCCC(C)CCCC(C)CCCC(C)C)C)C
Similarity: 0.621
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 103-23-1
Dataset id:315 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.323 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.65 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✗	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..
✓	Model class assignment reliability Pos/Non-Pos difference = 0.267 Explanation: model class assignment is well defined..
✓	Neural map neurons concordance Neurons concordance = 1 Explanation: predicted value agrees with experimental values of training set compounds laying in the same neuron..

Symbols explanation:

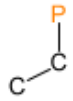
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

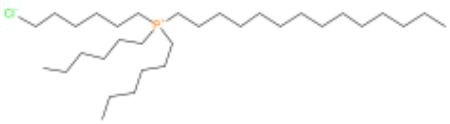
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

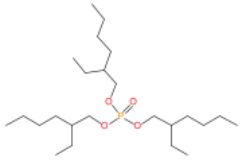


Prediction: Reliability:

Prediction is Carcinogen, but the result shows some critical aspects, which require to be checked:
- Only moderately similar compounds with known experimental value in the training set have been found
- 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 alerts have been found: SA8 Aliphatic halogens

Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Carcinogen activity: Carcinogen
Structural Alerts: SA8 Aliphatic halogens
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

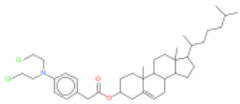
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 78-42-2
Dataset id:69 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : Carcinogen
Predicted value : Carcinogen

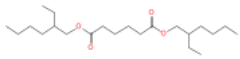
Alerts (not found also in the target): SA41 Substituted n-alkylcarboxylic acids



Compound #2

CAS: 3546-10-9
Dataset id:216 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : Carcinogen
Predicted value : Carcinogen

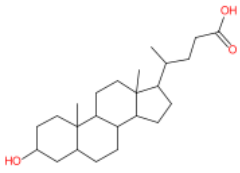
Alerts (not found also in the target): SA5 S or N mustard



Compound #3

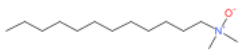
CAS: 103-23-1
Dataset id:52 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): SA41 Substituted n-alkylcarboxylic acids; SA42 Phthalate diesters and monoesters



Compound #4

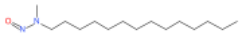
CAS: 434-13-9
Dataset id:117 (Training Set)
SMILES: O=C(O)CCC(C)C2CCC3C4CCC1CC(O)CCC1(C)C4(CCC23(C))
Similarity: 0.607
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 1643-20-5
Dataset id:879 (Training Set)
SMILES: [O-][N+](C)(C)CCCCCCCCCCCC
Similarity: 0.597
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

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



Compound #6

CAS: 75881-20-8
Dataset id:579 (Training Set)
SMILES: O=NN(C)CCCCCCCCCCCCCC
Similarity: 0.595
Experimental value : Carcinogen
Predicted value : Carcinogen

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

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.678 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.637 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
⚠	Atom Centered Fragments similarity check ACF index = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)..

Symbols explanation:

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

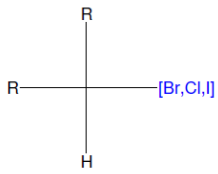
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

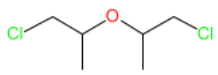
Fragment found: SA8 Aliphatic halogens



R = any atom/group

This alert contains non tertiary aliphatic halogens. Substances fired by Alerts SA2, SA4, SA5 and SA20 should be also excluded.

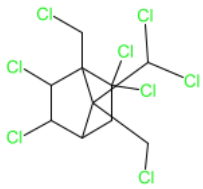
Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 108-60-1
Dataset id:40 (Training Set)
SMILES: O(C(C)CCl)C(C)CCl
Similarity: 0.526

Experimental value : Carcinogen
Predicted value : Carcinogen

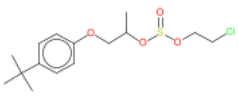
Alerts (found also in the target): SA8 Aliphatic halogens



CAS: 8001-35-2
Dataset id:727 (Training Set)
SMILES: C2C1C(C(C(CCl)(C1(CCl)C(Cl)Cl)C2(Cl)Cl)Cl)Cl
Similarity: 0.522

Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): SA8 Aliphatic halogens



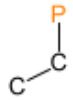
CAS: 140-57-8
Dataset id:478 (Training Set)
SMILES: O=S(OCCCl)OC(C)COc1ccc(cc1)C(C)(C)C
Similarity: 0.509

Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): SA8 Aliphatic halogens

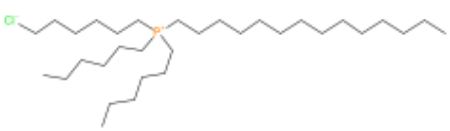
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

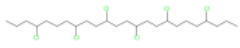


Prediction: Reliability:

Prediction is Possible NON-Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- similar molecules found in the training set have experimental values that disagree with the predicted value
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)

Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
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

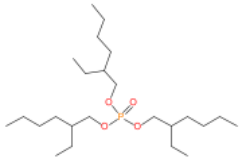
CAS: 63449-39-8
Dataset id:810 (Training Set)
SMILES: CCCC(CCCC(CCC(CCC(CCC(CCCC(CCC)Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.742
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 18

Compound #2

CAS: 108171-27-3
Dataset id:675 (Training Set)
SMILES: CC(CCCC(CCCC(CCCC(CCCC(CCCC(CCCl)Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.735
Experimental value : Carcinogen
Predicted value : Carcinogen

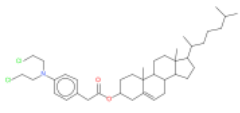
Alerts (not found also in the target): Carcinogenity alert no. 18



Compound #3

CAS: 78-42-2
Dataset id:59 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : Carcinogen
Predicted value : Carcinogen

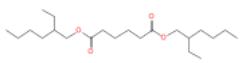
Alerts (not found also in the target): Carcinogenity alert no. 15



Compound #4

CAS: 3546-10-9
Dataset id:176 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 5; Carcinogenity alert no. 11; Carcinogenity alert no. 38; Carcinogenity alert no. 39; Carcinogenity alert no. 41; Carcinogenity alert no. 42

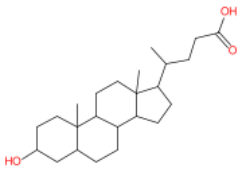


Compound #5

CAS: 103-23-1
Dataset id:43 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 29

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



Compound #6

CAS: 434-13-9
Dataset id:93 (Training Set)
SMILES: O=C(O)CCC(C)C2CCC3C4CCC1CC(O)CCC1(C)C4(CCC23(C))
Similarity: 0.607
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.704 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 0 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Atom Centered Fragments similarity check ACF index = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)..

Symbols explanation:

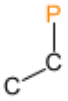
✔ The feature has a good assessment, model is reliable regarding this aspect.

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

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

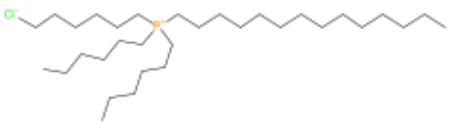
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -

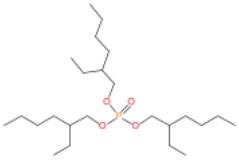


Prediction: Reliability:

Prediction is 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 found)
The following relevant fragments have been found: Carcinogenity alert no. 57

Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Carcinogenic activity: Carcinogen
No. alerts for carcinogenicity: 1
Structural Alerts: Carcinogenity alert no. 57
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

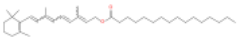
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

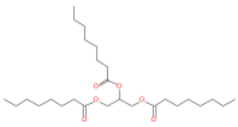
CAS: 78-42-2
Dataset id:784 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

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



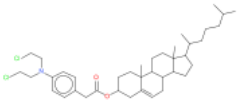
Compound #2

CAS: 79-81-2
Dataset id:693 (Training Set)
SMILES: O=C(OCC=C(C=CC=C(C=CC1=C(C)CCCC1(C)C)C)C)CCCCCCCCCCCCCCCC
Similarity: 0.649
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #3

CAS: 538-23-8
Dataset id:759 (Test Set)
SMILES: O=C(OCC(OC(=O)CCCCCCC)COC(=O)CCCCCCC)CCCCCCC
Similarity: 0.631
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen

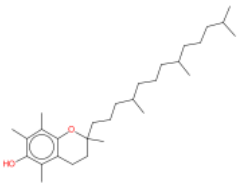


Compound #4

CAS: 3546-10-9
Dataset id:636 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 57

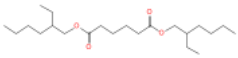
Alerts (not found also in the target): Carcinogenity alert no. 31; Carcinogenity alert no. 72; Carcinogenity alert no. 73; Carcinogenity alert no. 85



Compound #5

CAS: 10191-41-0
Dataset id:749 (Training Set)
SMILES: Oc2c(c(c1OC(C)(CCc1c2C)CCCC(C)CCCC(C)CCCC(C)C)C)C
Similarity: 0.621
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen

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



Compound #6

CAS: 103-23-1
Dataset id:315 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.184 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.644 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 0.338 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✖	Concordance for similar molecules Concordance index = 0.321 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✖	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

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

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

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

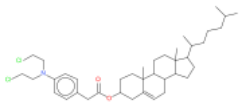
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

Fragment found: Carcinogenity alert no. 57

Structural alert for carcinogenity defined by the SMARTS: CCl

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

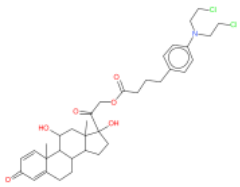


CAS: 3546-10-9
Dataset id:636 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623

Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 57

Alerts (not found also in the target): Carcinogenity alert no. 31; Carcinogenity alert no. 72; Carcinogenity alert no. 73; Carcinogenity alert no. 85

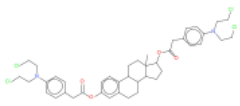


CAS: 29069-24-7
Dataset id:667 (Training Set)
SMILES:
O=C1C=CC4(C(=C1)CCC3C5CCC(O)(C(=O)COC(=O)CCCc2ccc(cc2)N(CCCl)CCCl)C5(C)(CC(O)C34))(C)
Similarity: 0.557

Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 57

Alerts (not found also in the target): Carcinogenity alert no. 31; Carcinogenity alert no. 72; Carcinogenity alert no. 73; Carcinogenity alert no. 86; Carcinogenity alert no. 106



CAS: 22966-79-6
Dataset id:297 (Training Set)
SMILES:
O=C(Oc1ccc2c(c1)CCC4C2CCC5(C)(C(OC(=O)Cc3ccc(cc3)N(CCCl)CCCl)CCC45))Cc6ccc(cc6)N(CCCl)CCCl
Similarity: 0.544

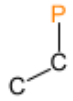
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 57

Alerts (not found also in the target): Carcinogenity alert no. 31; Carcinogenity alert no. 72; Carcinogenity alert no. 73; Carcinogenity alert no. 102

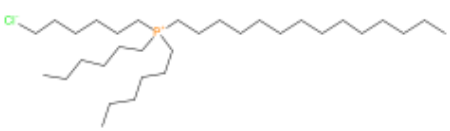
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

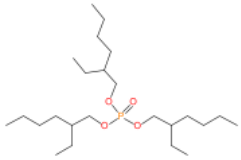


Prediction: Reliability:

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
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)

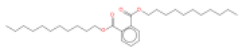
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Oral Carcinogenic class: Carcinogen
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



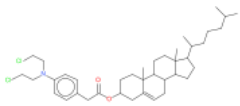
Compound #1

CAS: 78-42-2
Dataset id:313 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : Carcinogen
Predicted value : Carcinogen



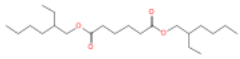
Compound #2

CAS: 3648-20-2
Dataset id:488 (Training Set)
SMILES: O=C(OCCCCCCCCCCC)c1ccccc1(C(=O)OCCCCCCCCCCC)
Similarity: 0.63
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



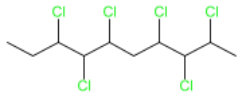
Compound #3

CAS: 3546-10-9
Dataset id:256 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : Carcinogen
Predicted value : Carcinogen



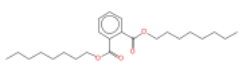
Compound #4

CAS: 103-23-1
Dataset id:94 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 108171-26-2
Dataset id:65 (Training Set)
SMILES: CCC(C(C(CC(C(C(C)Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.585
Experimental value : Carcinogen
Predicted value : Carcinogen



Compound #6

CAS: 117-84-0
Dataset id:614 (Training Set)
SMILES: O=C(OCCCCCCCC)c1ccccc1(C(=O)OCCCCCCCC)
Similarity: 0.581
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.407 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.641 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 0.515 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✖	Atom Centered Fragments similarity check ACF index = 0.6 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)..

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.

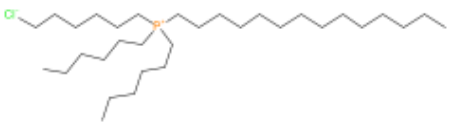
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(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: CP(C)(C)C
The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

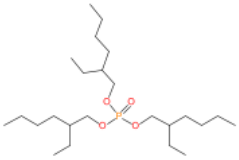


Prediction: Reliability:

Prediction is -1.09, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)

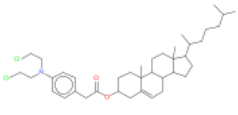
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Oral Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: -1.09
Predicted Oral Carcinogenicity SF [1/(mg/kg-day)]: 0.0817
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



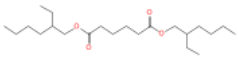
Compound #1

CAS: 78-42-2
Dataset id:313 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : -2.49
Predicted value : -2.174



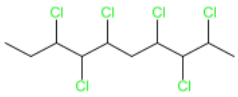
Compound #2

CAS: 3546-10-9
Dataset id:256 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : 2.18
Predicted value : 0.78



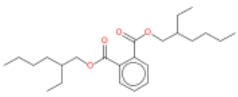
Compound #3

CAS: 103-23-1
Dataset id:94 (Test Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : -2.92
Predicted value : -1.999



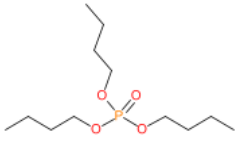
Compound #4

CAS: 108171-26-2
Dataset id:65 (Training Set)
SMILES: CCC(C(C(CC(C(C(C)Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.585
Experimental value : -1.05
Predicted value : -1.089



Compound #5

CAS: 117-81-7
Dataset id:44 (Test Set)
SMILES: O=C(OCC(CC)CCCC)c1ccccc1(C(=O)OCC(CC)CCCC)
Similarity: 0.566
Experimental value : -1.85
Predicted value : -2.919



Compound #6

CAS: 126-73-8
Dataset id:299 (Training Set)
SMILES: O=P(OCCCC)(OCCCC)OCCCC
Similarity: 0.545
Experimental value : -2.05
Predicted value : -2.145

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.255 Explanation: The predicted compound is outside the Applicability Domain of the model.
✖	Similar molecules with known experimental value Similarity index = 0.637 Explanation: No similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.858 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✖	Concordance for similar molecules Concordance index = 2.335 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Maximum error of prediction among similar molecules Max error index = 1.4 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✖	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

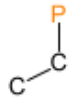
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

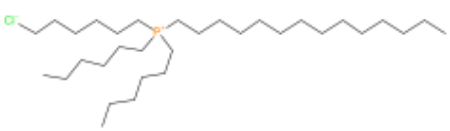
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

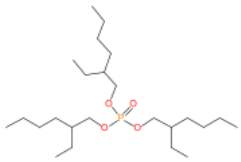


Prediction: Reliability:

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
- similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)

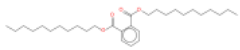
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Inhalation Carcinogenic class: Carcinogen
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



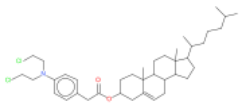
Compound #1

CAS: 78-42-2
Dataset id:741 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



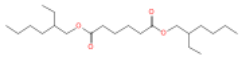
Compound #2

CAS: 3648-20-2
Dataset id:460 (Test Set)
SMILES: O=C(OCCCCCCCCCCC)c1ccccc1(C(=O)OCCCCCCCCCCC)
Similarity: 0.63
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



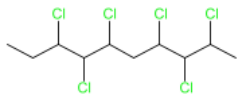
Compound #3

CAS: 3546-10-9
Dataset id:219 (Training Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : Carcinogen
Predicted value : Carcinogen



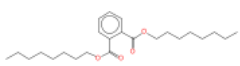
Compound #4

CAS: 103-23-1
Dataset id:391 (Training Set)
SMILES: O=C(OCC(CC)CCCC)CCCCC(=O)OCC(CC)CCCC
Similarity: 0.618
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 108171-26-2
Dataset id:54 (Training Set)
SMILES: CCC(C(C(CC(C(C(C)Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.585
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 117-84-0
Dataset id:597 (Training Set)
SMILES: O=C(OCCCCCCCC)c1ccccc1(C(=O)OCCCCCCCC)
Similarity: 0.581
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.641 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 0 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✖	Atom Centered Fragments similarity check ACF index = 0.6 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments found)..

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.

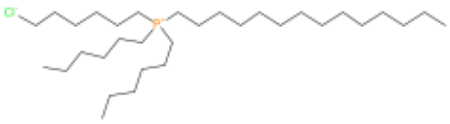
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(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: CP(C)(C)C
The fragment has never been found in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

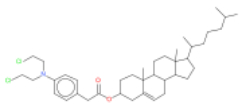


Prediction: Reliability:

Prediction is -0.06, 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 (2 unknown fragments found)

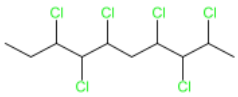
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted Inhalation Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: -0.06
Predicted Inhalation Carcinogenicity SF [1/(mg/kg-day)]: 0.8777
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



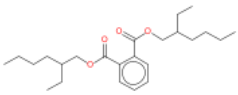
Compound #1

CAS: 3546-10-9
Dataset id:219 (Test Set)
SMILES:
O=C(OC4CC3=CCC1C(CCC2(C)(C(CCC12)C(C)CCCC(C)C))C3(C)CC4)Cc5ccc(cc5)N(CCCl)CCCl
Similarity: 0.623
Experimental value : 2.18
Predicted value : 1.825



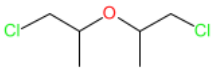
Compound #2

CAS: 108171-26-2
Dataset id:54 (Training Set)
SMILES: CCC(C(C(CC(C(C(C)Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.585
Experimental value : -1.06
Predicted value : 0.58



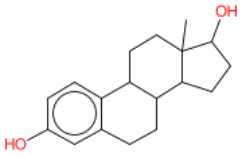
Compound #3

CAS: 117-81-7
Dataset id:38 (Training Set)
SMILES: O=C(OCC(CC)CCCC)c1ccccc1(C(=O)OCC(CC)CCCC)
Similarity: 0.566
Experimental value : -2.08
Predicted value : 0.163



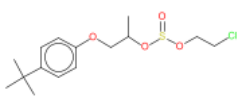
Compound #4

CAS: 108-60-1
Dataset id:36 (Training Set)
SMILES: O(C(C)CCl)C(C)CCl
Similarity: 0.526
Experimental value : -1.46
Predicted value : 0.263



Compound #5

CAS: 50-28-2
Dataset id:116 (Training Set)
SMILES: Oc1ccc2c(c1)CCC3C2CCC4(C)(C(O)CCC34)
Similarity: 0.521
Experimental value : 1.59
Predicted value : 2.033



Compound #6

CAS: 140-57-8
Dataset id:237 (Training Set)
SMILES: O=S(OCCCl)OC(C)COc1ccc(cc1)C(C)(C)C
Similarity: 0.509
Experimental value : -1.6
Predicted value : -0.214

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.241 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.603 Explanation: No similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.998 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✗	Concordance for similar molecules Concordance index = 1.62 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✗	Maximum error of prediction among similar molecules Max error index = 1.64 Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✗	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

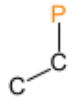
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

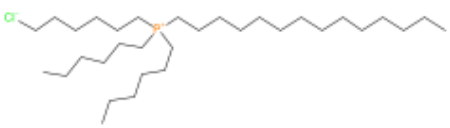
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

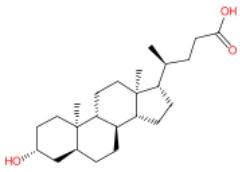


Prediction: Reliability:

Prediction is -3.3514, 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
- 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 (2 unknown fragments found)

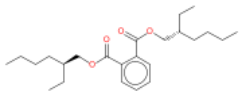
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: -3.3514
Predicted TD50 [mg/kg bw/day]: 2245.76
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



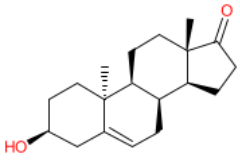
Compound #1

CAS: N.A.
Dataset id:129 (Training Set)
SMILES:
[C@@H]1(C[C@H]2[C@@](CC1)([C@H]1[C@H](CC2)[C@@H]2[C@@](CC1)([C@@H](C
C2)[C@@H](C)CCC(=O)O)C)C)O
Similarity: 0.607
Experimental value : -3.352
Predicted value : -0.763



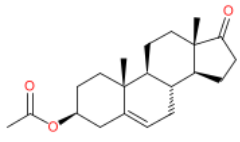
Compound #2

CAS: N.A.
Dataset id:122 (Training Set)
SMILES: c1ccc(c(c1)C(=O)OC[C@H](CCCC)CC)C(=O)OC[C@@H](CCCC)CC
Similarity: 0.566
Experimental value : -3.068
Predicted value : -2.992



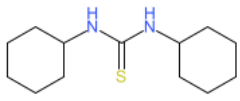
Compound #3

CAS: N.A.
Dataset id:164 (Test Set)
SMILES:
C12=CC[C@H]3[C@@H]([C@@]2(CC[C@@H](C1)O)C)CC[C@]1([C@H]3CCC1=O)C
Similarity: 0.562
Experimental value : -1.833
Predicted value : -0.788



Compound #4

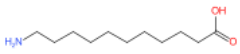
CAS: N.A.
Dataset id:81 (Training Set)
SMILES:
[C@H]1(CC2=CC[C@@H]3[C@@H]([C@]2(CC1)C)CC[C@]1([C@H]3CCC1=O)C)OC(=O)
C
Similarity: 0.559
Experimental value : -1.559
Predicted value : -1.497



Compound #5

CAS: N.A.
Dataset id:176 (Test Set)
SMILES: C1C(CCCC1)NC(=S)NC1CCCCC1
Similarity: 0.554
Experimental value : -4.193
Predicted value : -3.794

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



Compound #6

CAS: N.A.
Dataset id:34 (Training Set)
SMILES: C(CCCCCCCCCC(=O)O)N
Similarity: 0.548
Experimental value : -3.041
Predicted value : -2.357

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.234 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.586 Explanation: No similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 1.333 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✓	Concordance for similar molecules Concordance index = 0.142 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✗	Maximum error of prediction among similar molecules Max error index = 2.589 Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✗	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

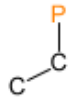
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

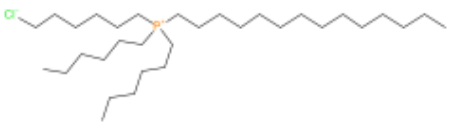
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

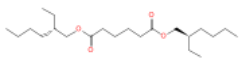


Prediction: Reliability:

Prediction is -1.6429, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)

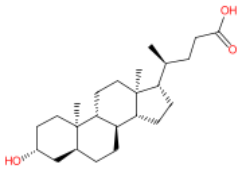
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: -1.6429
Predicted TD50 [mg/kg bw/day]: 43.94
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



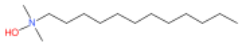
Compound #1

CAS: N.A.
Dataset id:113 (Training Set)
SMILES: CCCC[C@H](CC)COC(=O)CCCCC(=O)OC[C@H](CC)CCCC
Similarity: 0.618
Experimental value : -4.161
Predicted value : -2.71



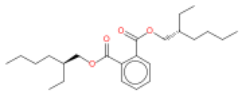
Compound #2

CAS: N.A.
Dataset id:47 (Training Set)
SMILES:
[C@@H]1(C[C@H]2[C@@](CC1)([C@H]1[C@H](CC2)[C@@H]2[C@@](CC1)([C@@H](CC2)[C@@H](C)CCC(=O)O)C)C)O
Similarity: 0.607
Experimental value : -3.276
Predicted value : -4.644



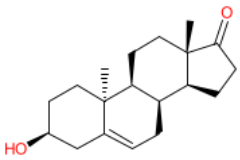
Compound #3

CAS: N.A.
Dataset id:93 (Training Set)
SMILES: CCCCCCCCCCCC[N](O)(C)C
Similarity: 0.597
Experimental value : -2.364
Predicted value : -0.843



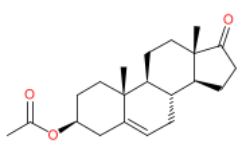
Compound #4

CAS: N.A.
Dataset id:42 (Training Set)
SMILES: c1ccc(c(c1)C(=O)OC[C@H](CCCC)CC)C(=O)OC[C@@H](CCCC)CC
Similarity: 0.566
Experimental value : -3.045
Predicted value : -2.938



Compound #5

CAS: N.A.
Dataset id:25 (Training Set)
SMILES:
C12=CC[C@H]3[C@@H]([C@@]2(CC[C@@H](C1)O)C)CC[C@]1([C@H]3CCC1=O)C
Similarity: 0.562
Experimental value : -1.921
Predicted value : -4.971



Compound #6

CAS: N.A.
Dataset id:33 (Training Set)
SMILES:
[C@H]1(CC2=CC[C@@H]3[C@@H]([C@]2(CC1)C)CC[C@]1([C@H]3CCC1=O)C)OC(=O)C
Similarity: 0.559
Experimental value : -2.362
Predicted value : -1.307

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.245 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.612 Explanation: No similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 1.41 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✗	Concordance for similar molecules Concordance index = 2.076 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Maximum error of prediction among similar molecules Max error index = 1.451 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✗	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

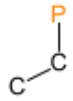
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

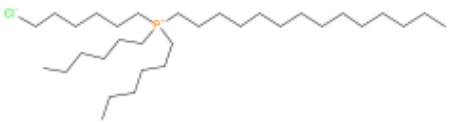
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

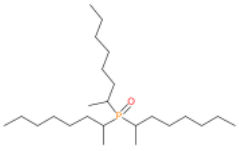


Prediction: Reliability:

Prediction is 10992.87 mg/kg, the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections.

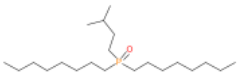
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted log LD50 [log(mmol/Kg)]: 1.326
Predicted log LD50 [mg/Kg]: 10992.87
Molecules used for prediction: 2
Experimental value [mg/Kg]: -
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



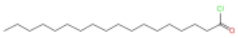
Compound #1

CAS: N.A.
Dataset id:2782 (Training Set)
SMILES: O=P(C(C)CCCCCC)(C(C)CCCCCC)C(C)CCCCCC
Similarity: 0.841
Experimental value : 1.27
Predicted value : 1.324



Compound #2

CAS: N.A.
Dataset id:2575 (Training Set)
SMILES: O=P(CCCCCCCC)(CCCCCCCC)CCC(C)C
Similarity: 0.808
Experimental value : 1.39
Predicted value : 1.266

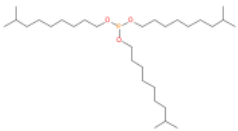


Compound #3

CAS: N.A.
Dataset id:5570 (Training Set)
SMILES: O=C(CCCCCCCCCCCCCCCCC)Cl
Similarity: 0.741
Experimental value : 1.39
Predicted value : 1.494

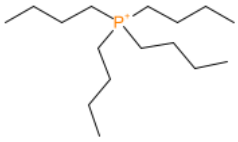
Compound #4

CAS: N.A.
Dataset id:5710 (Training Set)
SMILES: OC(C[N+](C)(C)CCCCCCCCCCCCCCCCCC)CCl
Similarity: 0.722
Experimental value : 0.78
Predicted value : 0.202



Compound #5

CAS: N.A.
Dataset id:5782 (Training Set)
SMILES: O(CCCCCCCC(C)C)P(OCCCCCCCC(C)C)OCCCCCCCC(C)C
Similarity: 0.721
Experimental value : 1.38
Predicted value : 1.264



Compound #6

CAS: N.A.
Dataset id:1894 (Training Set)
SMILES: CCCC[P+](CCCC)(CCCC)CCCC
Similarity: 0.72
Experimental value : 0.09
Predicted value : 0.902

3.2 Applicability Domain:
Measured Applicability Domain Scores

✓	Global AD Index AD index = 1 Explanation: The predicted compound is into the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.824 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.089 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0.06 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✓	Maximum error of prediction among similar molecules Max error index = 0.124 Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability..
✓	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

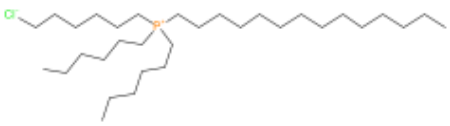
✓ The feature has a good assessment, model is reliable regarding this aspect.

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -



Prediction: Reliability:

Prediction is 0.63 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
- 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 moderate value, considering the experimental variability
- 2descriptor(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 found)

Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted BCF [log(L/kg)]: 0.63
Predicted BCF [L/kg]: 4
Predicted BCF from sub-model 1 (HM) [log(L/kg)]: 1.47
Predicted BCF from sub-model 2 (GA) [log(L/kg)]: 0.13
Predicted LogP (MLogP): 10.01
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

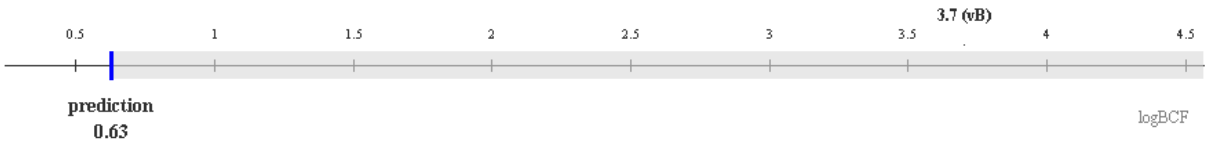
2. Possible Use and Uncertainty

Threshold 3.3 (bioaccumulative)

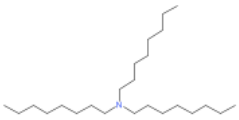
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.3, the current compound can not be associated (due to its Applicability Domain index value) to any conservative interval.
No safe classification can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



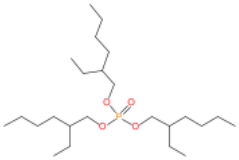
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 1116-76-3
Dataset id:306 (Training Set)
SMILES: N(CCCCCCCC)(CCCCCCCC)CCCCCCCC
Similarity: 0.66
Experimental value : 1.92
Predicted value : 1.35

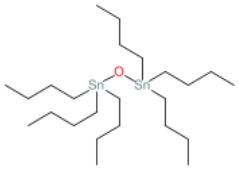
Alerts (not found also in the target): Tertiary amine (SR 05)



Compound #2

CAS: 78-42-2
Dataset id:405 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : 1.19
Predicted value : 1.31

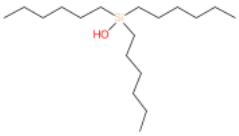
Alerts (not found also in the target): PO2 residue (SR 03)



Compound #3

CAS: 56-35-9
Dataset id:466 (Training Set)
SMILES: O([Sn](CCCC)(CCCC)CCCC)[Sn](CCCC)(CCCC)CCCC
Similarity: 0.651
Experimental value : 3.85
Predicted value : 3.686

Alerts (not found also in the target): Sn atom in the molecule (SO 04)



Compound #4

CAS: 60782-58-3
Dataset id:465 (Training Set)
SMILES: O[Si](CCCCCC)(CCCCCC)CCCCCC
Similarity: 0.635
Experimental value : 1.48
Predicted value : 1.927

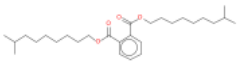
Alerts (not found also in the target): Si atom in the molecule (SO 03); OH group (PG 06)

Compound #5

CAS: 28299-29-8
Dataset id:290 (Training Set)
SMILES: O=C(O)CC(C(=O)O)CCCCCCCCC=CCCCCCCCC
Similarity: 0.623
Experimental value : 0.22
Predicted value : 0.993

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

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



Compound #6

CAS: 26761-40-0
Dataset id:302 (Training Set)
SMILES: O=C(OCCCCCCCC(C)C)c1ccccc1(C(=O)OCCCCCCCC(C)C)
Similarity: 0.605
Experimental value : 1.16
Predicted value : 0.77

Alerts (not found also in the target): Moiety (SMILES: O=Cc1ccccc1) (SR 01); Carbonyl residue (SR 02); >C=O group (PG 09)

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
✖	Similar molecules with known experimental value Similarity index = 0.656 Explanation: No similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.345 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
⚠	Concordance for similar molecules Concordance index = 0.929 Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Maximum error of prediction among similar molecules Max error index = 0.57 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability..
✖	Model's descriptors range check Descriptors range check = False Explanation: 2descriptor(s) for this compound have values outside the descriptor range of the compounds of the training set...
✖	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

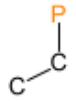
✔ The feature has a good assessment, model is reliable regarding this aspect.

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

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

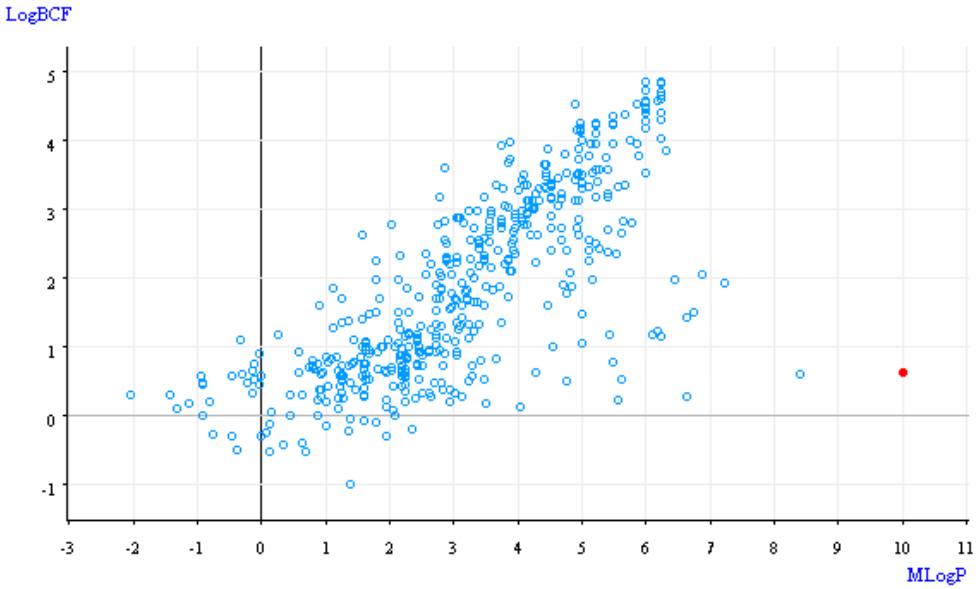
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

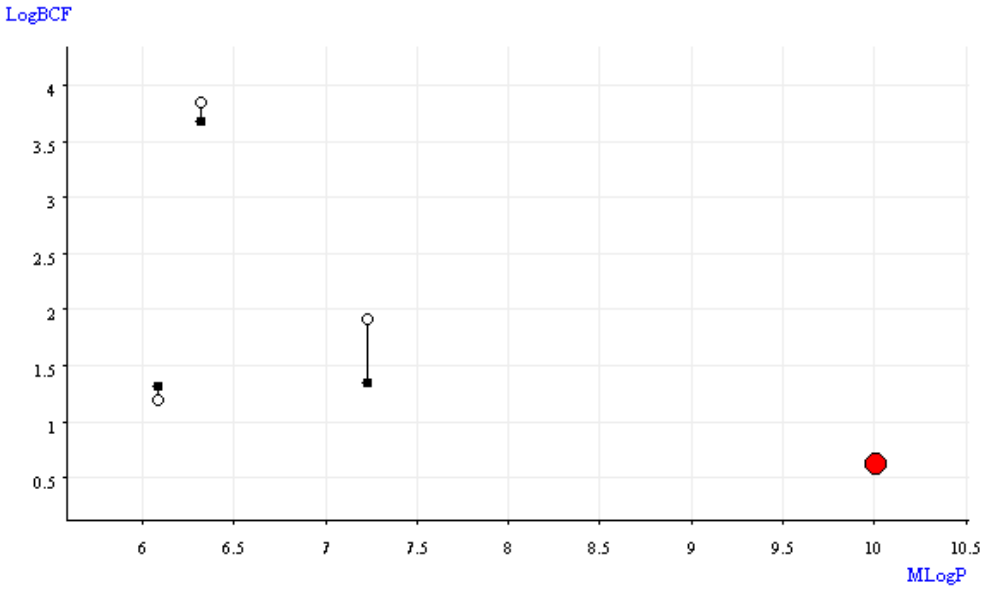
4.2 Reasoning:
Analysis of Molecular Descriptors

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.

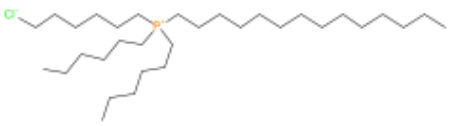


Following, a scatterplot of MLogP against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound;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



1. Prediction Summary

Prediction for compound Molecule 0 -

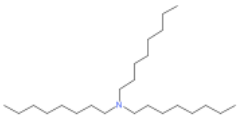


Prediction: Reliability:

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
- similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability
- reliability of logP value used by the model is not adequate
- LogP of this compound is outside the defined range [-1.37,11.26]
- 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)

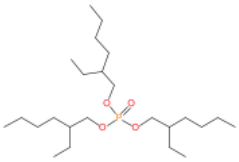
Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted BCF [log(L/kg)]: 0.5
Predicted BCF [L/kg]: 3
Predicted LogP (Meylan/Kowwin): 16.42
Predicted LogP reliability: Low
MW: 515.15
Ionic compound: no
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



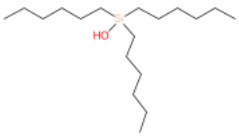
Compound #1

CAS: 1116-76-3
Dataset id:647 (Test Set)
SMILES: N(CCCCCCCC)(CCCCCCCC)CCCCCCCC
Similarity: 0.66
Experimental value : 1.93
Predicted value : 2.482



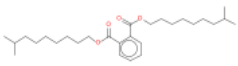
Compound #2

CAS: 78-42-2
Dataset id:522 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : 1.7
Predicted value : 1.482



Compound #3

CAS: 60782-58-3
Dataset id:518 (Training Set)
SMILES: O[Si](CCCCCC)(CCCCCC)CCCCCC
Similarity: 0.635
Experimental value : 2.7
Predicted value : 2.098

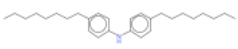


Compound #4

CAS: 26761-40-0
Dataset id:525 (Training Set)
SMILES: O=C(OCCCCCCCCC(C)C)c1ccccc1(C(=O)OCCCCCCCCC(C)C)
Similarity: 0.605
Experimental value : 1.16
Predicted value : 2.479

Compound #5

CAS: 544-76-3
Dataset id:515 (Training Set)
SMILES: CCCCCCCCCCCCCCCC
Similarity: 0.602
Experimental value : 3.7
Predicted value : 2.94



Compound #6

CAS: 26603-23-6
Dataset id:526 (Training Set)
SMILES: c1cc(ccc1Nc2ccc(cc2)CCCCCCCC)CCCCCCCC
Similarity: 0.599
Experimental value : 1.54
Predicted value : 2.036

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
✖	Similar molecules with known experimental value Similarity index = 0.656 Explanation: No similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.385 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✖	Concordance for similar molecules Concordance index = 1.315 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
⚠	Maximum error of prediction among similar molecules Max error index = 0.552 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability..
✖	Reliability of logP prediction LogP reliability = 0 Explanation: reliability of logP value used by the model is not adequate..
✖	Model's descriptors range check Descriptors range check = False Explanation: LogP of this compound is outside the defined range [-1.37,11.26]..
✖	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

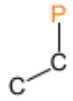
✔ The feature has a good assessment, model is reliable regarding this aspect.

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

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

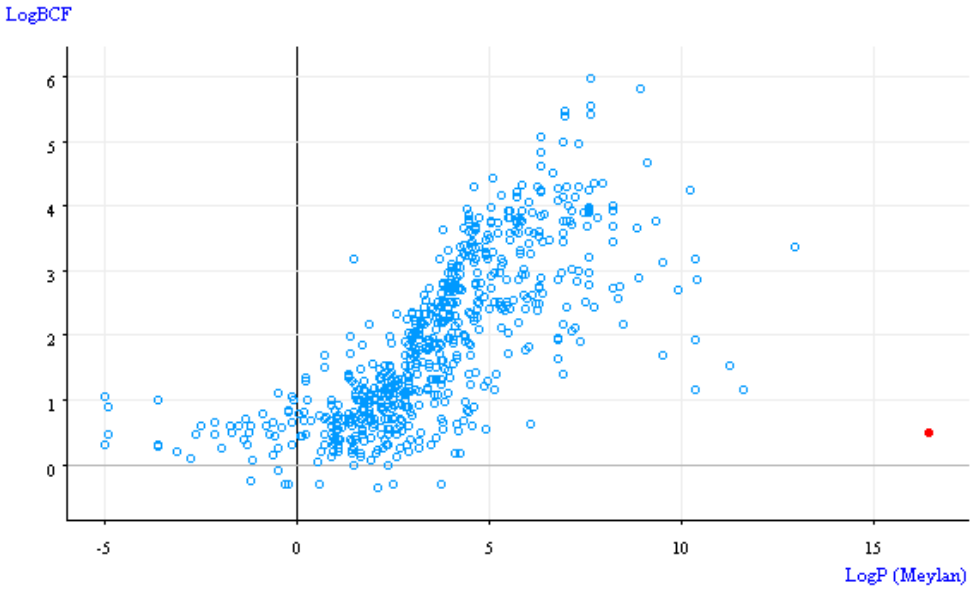
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

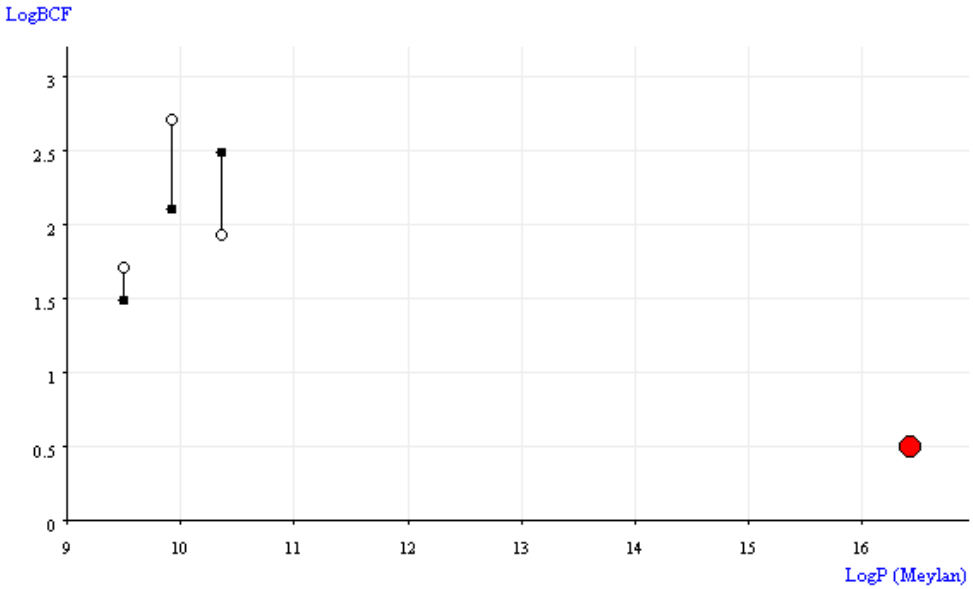
4.2 Reasoning:
Analysis of Molecular Descriptors

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.

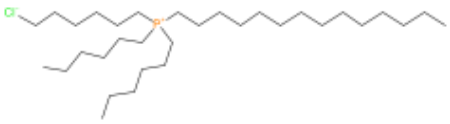


Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same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

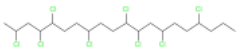
Prediction is -0.05 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate
- 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
- 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: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted BCF (up) [log(L/kg)]: -0.05
Predicted BCF (up) [L/kg]: 0.89
Predicted BCF (low) [log(L/kg)]: -0.03
Predicted BCF (low) [L/kg]: 0.94
Predicted BCF (mid) [log(L/kg)]: -0.03
Predicted BCF (mid) [L/kg]: 0.93
Predicted LogP (Meylan/Kowwin): 16.42
Predicted LogP reliability: Low
Predicted kM (Meylan): 4.08
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: 4051-66-5
Dataset id:780 (Training Set)
SMILES: O=C(NCCCCCCCCCCCCCCCCCC)NCCCCCCCCCCCCCCCCCC
Similarity: 0.691
Experimental value : 1.235
Predicted value : -0.049



Compound #2

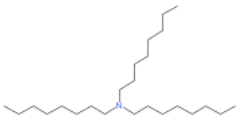
CAS: 61788-76-9
Dataset id:433 (Training Set)
SMILES: CCCC(CCC(CC(C(CCC(CCC(C(CC(C)Cl)Cl)Cl)Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.685
Experimental value : 1.69
Predicted value : 0.304

Compound #3

CAS: 10496-18-1
Dataset id:99 (Training Set)
SMILES: CCCCCCCCCCSSCCCCCCCCCC
Similarity: 0.667
Experimental value : 1.15
Predicted value : 0.516

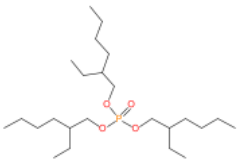
Compound #4

CAS: 110-30-5
Dataset id:315 (Training Set)
SMILES: O=C(NCCNC(=O)CCCCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCCCCC
Similarity: 0.666
Experimental value : 0.32
Predicted value : -0.049



Compound #5

CAS: 1116-76-3
Dataset id:72 (Training Set)
SMILES: N(CCCCCCCC)(CCCCCCCC)CCCCCCCC
Similarity: 0.66
Experimental value : 1.86
Predicted value : 0.412



Compound #6

CAS: 78-42-2
Dataset id:199 (Training Set)
SMILES: O=P(OCC(CC)CCCC)(OCC(CC)CCCC)OCC(CC)CCCC
Similarity: 0.652
Experimental value : 0.865
Predicted value : 0.087

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.275 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.688 Explanation: No similar compounds with known experimental value in the training set have been found..
✗	Accuracy of prediction for similar molecules Accuracy index = 1.335 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✗	Concordance for similar molecules Concordance index = 1.512 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✗	Maximum error of prediction among similar molecules Max error index = 1.386 Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability..
✗	Reliability of logP prediction LogP reliability = 0 Explanation: reliability of logP value used by the model is not adequate..
✗	Atom Centered Fragments similarity check ACF index = 0.4 Explanation: a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 unknown fragments found)..

Symbols explanation:

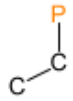
✓ The feature has a good assessment, model is reliable regarding this aspect.

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

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

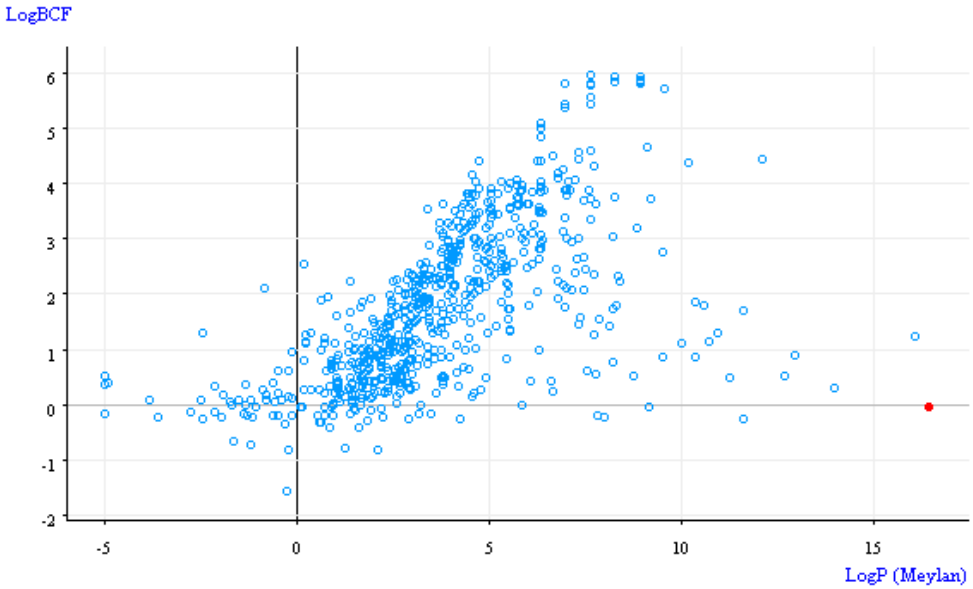
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

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

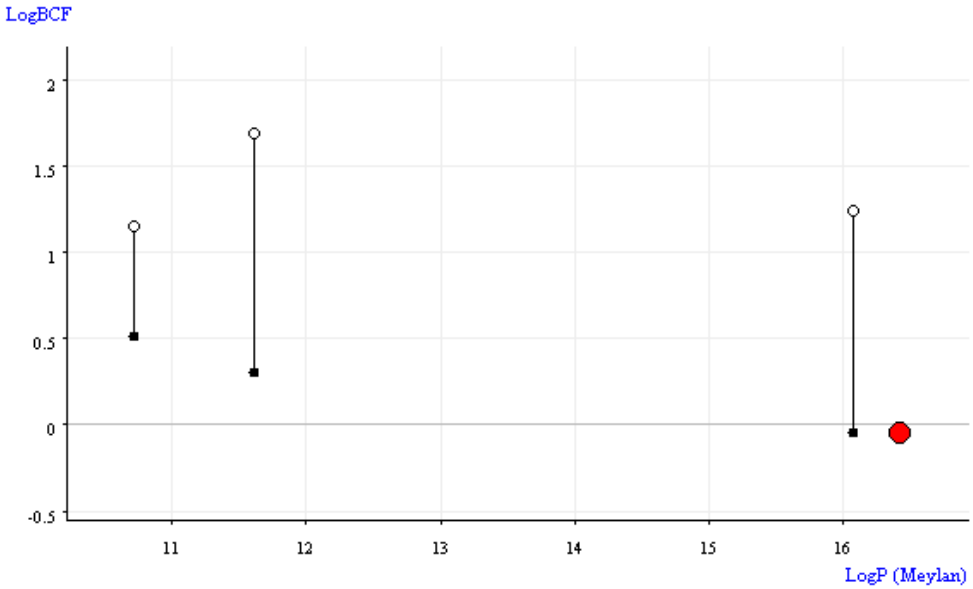
4.2 Reasoning:
Analysis of Molecular Descriptors

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.

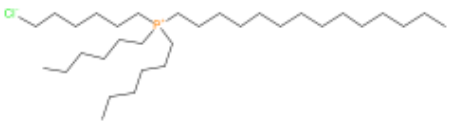


Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same 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 N/A, it is not possible to perform an assessment.

Compound: Molecule 0
Compound SMILES: CCCCCCCCCCCCCC[P+](CCCCCC)(CCCCCC)CCCCCC[Cl-]
Experimental value: -
Predicted BCF [log(L/kg)]: -
Molecules used for prediction: 0
Reliability: -
Remarks:
[Model] Unable to perform Applicability Domain check

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