VEGA QSAR MODELS

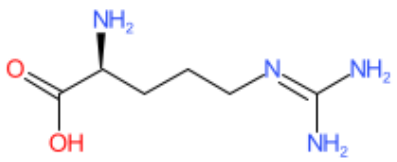
Report

Prediction and Applicability Domain analysis for models:

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

Core version: 1.3.18

Prediction for compound Molecule 0 -



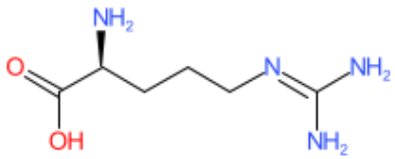
Prediction:

Prediction is NON-Mutagenic with a consensus score of 0.6, based on 4 models.

Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Used models: 4
Predicted Consensus Mutagen activity: NON-Mutagenic
Mutagenic Score: 0.15
Non-Mutagenic Score: 0.6
Model Caesar assessment: NON-Mutagenic (GOOD reliability)
Model ISS assessment: NON-Mutagenic (MODERATE reliability)
Model SarPy assessment: Possible NON-Mutagenic (GOOD reliability)
Model KNN assessment: Mutagenic (MODERATE reliability)
Remarks:
none

1. Prediction Summary

Prediction for compound Molecule 0 -

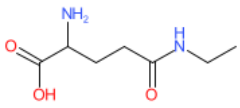


Prediction: Reliability:

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

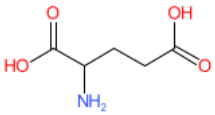
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



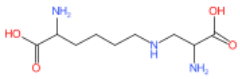
Compound #1

CAS: 3081-61-6
Dataset id:4237 (Training Set)
SMILES: O=C(O)C(N)CCC(=O)NCC
Similarity: 0.844
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



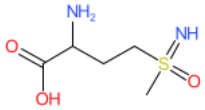
Compound #2

CAS: 56-86-0
Dataset id:4115 (Training Set)
SMILES: O=C(O)CCC(N)C(=O)O
Similarity: 0.809
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



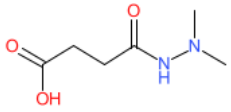
Compound #3

CAS: 18810-04-3
Dataset id:1086 (Test Set)
SMILES: O=C(O)C(N)CNCCCCC(N)C(=O)O
Similarity: 0.801
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

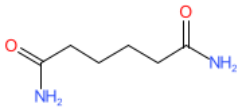
CAS: 1982-67-8
Dataset id:309 (Training Set)
SMILES: O=C(O)C(N)CCS(=O)(=N)C
Similarity: 0.796
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #5

CAS: 1596-84-5
Dataset id:1694 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

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



Compound #6

CAS: 628-94-4
Dataset id:334 (Training Set)
SMILES: O=C(N)CCCCC(=O)N
Similarity: 0.787
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

✓	Global AD Index AD index = 0.904 Explanation: The predicted compound is into the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.816 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✓	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

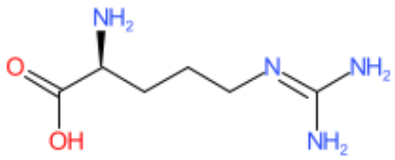
✓ The feature has a good assessment, model is reliable regarding this aspect.

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -

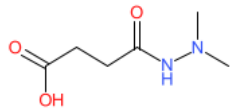


Prediction: Reliability:

Prediction is NON-Mutagenic, but the result shows some critical aspects, which require to be checked:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not adequate

Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

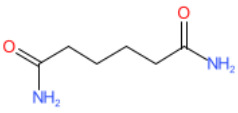
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

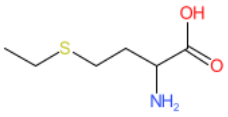
CAS: 1596-84-5
Dataset id:225 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

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



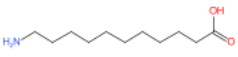
Compound #2

CAS: 628-94-4
Dataset id:782 (Training Set)
SMILES: O=C(N)CCCCC(=O)N
Similarity: 0.787
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



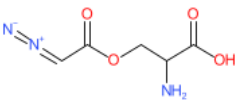
Compound #3

CAS: 13073-35-3
Dataset id:477 (Training Set)
SMILES: O=C(O)C(N)CCSCC
Similarity: 0.765
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #4

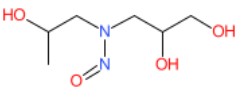
CAS: 2432-99-7
Dataset id:36 (Training Set)
SMILES: O=C(O)CCCCCCCCCCCN
Similarity: 0.75
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #5

CAS: 115-02-6
Dataset id:443 (Training Set)
SMILES: [N-]=[N+]=CC(=O)OCC(N)C(=O)O
Similarity: 0.743
Experimental value : Mutagenic
Predicted value : Mutagenic

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



Compound #6

CAS: 89911-79-5
Dataset id:578 (Training Set)
SMILES: O=NN(CC(O)C)CC(O)CO
Similarity: 0.74
Experimental value : Mutagenic
Predicted value : Mutagenic

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 = 0.746 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.788 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 0.499 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
✔	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✔	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

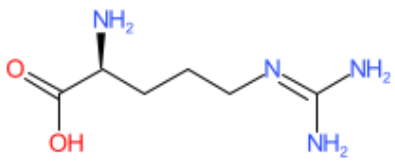
✔ The feature has a good assessment, model is reliable regarding this aspect.

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -

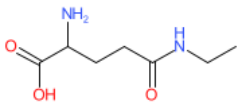


Prediction: Reliability:

Prediction is Possible NON-Mutagenic, the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections.

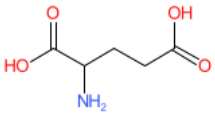
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Mutagen activity: Possible NON-Mutagenic
No. alerts for mutagenicity: 0
No. alerts for non-mutagenicity: 0
Structural Alerts: -
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



Compound #1

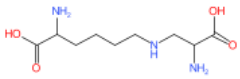
CAS: 3081-61-6
Dataset id:4237 (Training Set)
SMILES: O=C(O)C(N)CCC(=O)NCC
Similarity: 0.844
Experimental value : NON-Mutagenic
Predicted value : Possible NON-Mutagenic



Compound #2

CAS: 56-86-0
Dataset id:4115 (Training Set)
SMILES: O=C(O)CCC(N)C(=O)O
Similarity: 0.809
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

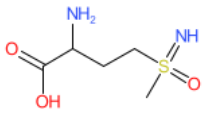
Alerts (not found also in the target): SM121



Compound #3

CAS: 18810-04-3
Dataset id:1086 (Test Set)
SMILES: O=C(O)C(N)CNCCCCC(N)C(=O)O
Similarity: 0.801
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

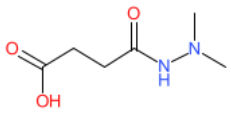
Alerts (not found also in the target): SM177



Compound #4

CAS: 1982-67-8
Dataset id:309 (Training Set)
SMILES: O=C(O)C(N)CCS(=O)(=N)C
Similarity: 0.796
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (not found also in the target): SM153

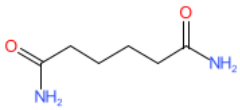


Compound #5

CAS: 1596-84-5
Dataset id:1694 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SM103; SM137

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



Compound #6

CAS: 628-94-4
Dataset id:334 (Training Set)
SMILES: O=C(N)CCCCC(=O)N
Similarity: 0.787
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Alerts (not found also in the target): SM177

3.2 Applicability Domain:
Measured Applicability Domain Scores

✓	Global AD Index AD index = 0.904 Explanation: The predicted compound is into the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.816 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✓	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 1 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
✓	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:

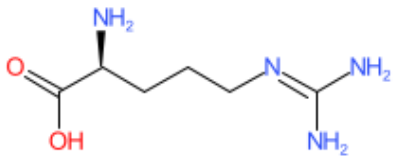
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 -

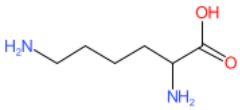


Prediction: Reliability:

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

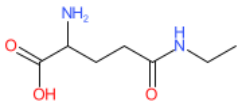
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Mutagen activity: Mutagenic
Molecules used for prediction: 4
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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



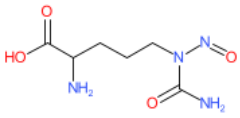
Compound #1

CAS: 70-54-2
Dataset id:4480 (Training Set)
SMILES: O=C(O)C(N)CCCCN
Similarity: 0.85
Experimental value : Mutagenic
Predicted value : NON-Mutagenic



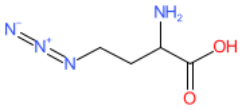
Compound #2

CAS: 3081-61-6
Dataset id:2672 (Training Set)
SMILES: O=C(O)C(N)CCC(=O)NCC
Similarity: 0.844
Experimental value : NON-Mutagenic
Predicted value : Mutagenic



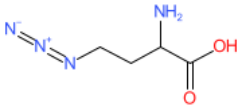
Compound #3

CAS: 42713-66-6
Dataset id:3077 (Training Set)
SMILES: O=NN(C(=O)N)CCCC(N)C(=O)O
Similarity: 0.818
Experimental value : Mutagenic
Predicted value : Mutagenic



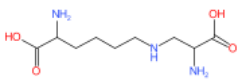
Compound #4

CAS: 120042-14-0
Dataset id:824 (Training Set)
SMILES: [N-]=[N+]=NCCC(N)C(=O)O
Similarity: 0.805
Experimental value : Mutagenic
Predicted value : Mutagenic



Compound #5

CAS: 120143-20-6
Dataset id:825 (Training Set)
SMILES: [N-]=[N+]=NCCC(N)C(=O)O
Similarity: 0.805
Experimental value : Mutagenic
Predicted value : Mutagenic



Compound #6

CAS: 18810-04-3
Dataset id:2014 (Training Set)
SMILES: O=C(O)C(N)CNCCCCC(N)C(=O)O
Similarity: 0.801
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.706 Explanation: The predicted compound could be out of the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.828 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✗	Accuracy of prediction for similar molecules Accuracy index = 0.487 Explanation: Accuracy of prediction for similar molecules found in the training set is not adequate..
⚠	Concordance for similar molecules Concordance index = 0.744 Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value..
✓	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

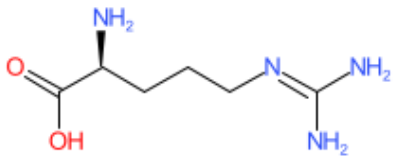
✓ The feature has a good assessment, model is reliable regarding this aspect.

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -

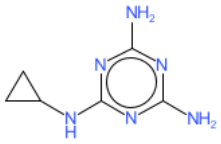


Prediction: Reliability:

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

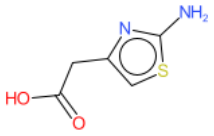
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Mutagen activity: NA
Structural Alerts: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



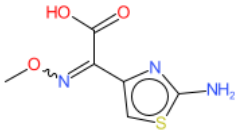
Compound #1

CAS: N.A.
Dataset id:6397 (Training Set)
SMILES: n1c(nc(nc1N)NC2CC2)N
Similarity: 0.701
Experimental value : NON-Mutagenic
Predicted value : NA



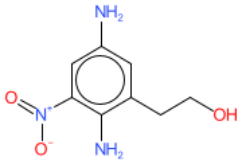
Compound #2

CAS: N.A.
Dataset id:7505 (Training Set)
SMILES: O=C(O)Cc1nc(N)sc1
Similarity: 0.693
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



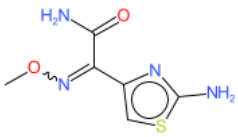
Compound #3

CAS: N.A.
Dataset id:5869 (Training Set)
SMILES: O=C(O)C(=NOC)c1nc(N)sc1
Similarity: 0.69
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



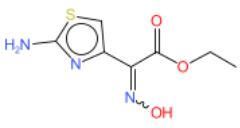
Compound #4

CAS: N.A.
Dataset id:1844 (Training Set)
SMILES: O=[N+]([O-])c1cc(N)cc(c1(N))CCO
Similarity: 0.688
Experimental value : Mutagenic
Predicted value : Mutagenic



Compound #5

CAS: N.A.
Dataset id:5366 (Training Set)
SMILES: O=C(N)C(=NOC)c1nc(N)sc1
Similarity: 0.684
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic



Compound #6

CAS: N.A.
Dataset id:4852 (Training Set)
SMILES: O=C(OCC)C(=NO)c1nc(N)sc1
Similarity: 0.679
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 = 0.697 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✖	Accuracy of prediction for similar molecules Accuracy index = 0.496 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:

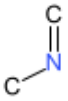
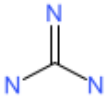
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

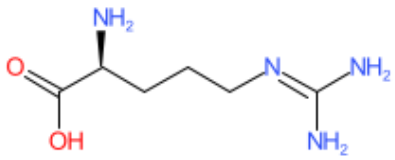
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on 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: CN=C The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: NC(N)=N The fragment has less than 3 occurrences in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

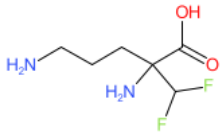


Prediction: Reliability:

Prediction is NON-Carcinogen, but the result shows some critical aspects, which require to be checked:
- some similar molecules found in the training set have experimental values that disagree with the predicted value

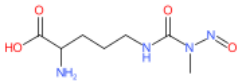
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Carcinogen activity: NON-Carcinogen
P(Carcinogen): 0.151
P(NON-Carcinogen): 0.849
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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



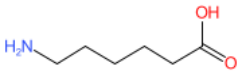
Compound #1

CAS: 70052-12-9
Dataset id:245 (Training Set)
SMILES: O=C(O)C(N)(CCCN)C(F)F
Similarity: 0.85
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



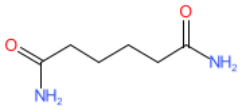
Compound #2

CAS: 63642-17-1
Dataset id:457 (Training Set)
SMILES: O=NN(C(=O)NCCCC(N)C(=O)O)C
Similarity: 0.805
Experimental value : Carcinogen
Predicted value : Carcinogen



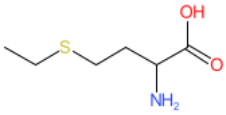
Compound #3

CAS: 60-32-2
Dataset id:47 (Training Set)
SMILES: O=C(O)CCCCCN
Similarity: 0.793
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



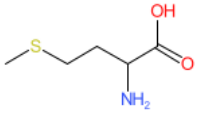
Compound #4

CAS: 628-94-4
Dataset id:23 (Training Set)
SMILES: O=C(N)CCCCC(=O)N
Similarity: 0.787
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 13073-35-3
Dataset id:299 (Test Set)
SMILES: O=C(O)C(N)CCSCC
Similarity: 0.765
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 59-51-8
Dataset id:435 (Training Set)
SMILES: O=C(O)C(N)CCSC
Similarity: 0.765
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.77 Explanation: The predicted compound could be out of the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.826 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✓	Accuracy of prediction for similar molecules Accuracy index = 1 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
⚠	Concordance for similar molecules Concordance index = 0.517 Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✓	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..
✓	Model class assignment reliability Pos/Non-Pos difference = 0.697 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:

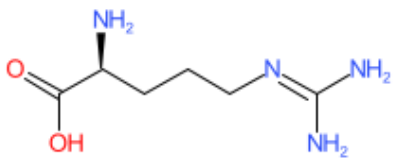
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

1. Prediction Summary

Prediction for compound Molecule 0 -

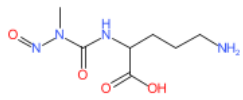


Prediction: Reliability:

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
- similar molecules found in the training set have experimental values that disagree with the predicted value

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

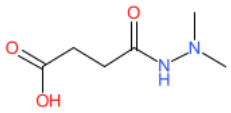
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 63642-17-1
Dataset id:757 (Training Set)
SMILES: O=NN(C(=O)NC(C(=O)O)CCCN)C
Similarity: 0.809
Experimental value : Carcinogen
Predicted value : Carcinogen

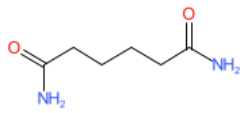
Alerts (not found also in the target): SA21 Alkyl and aryl N-nitroso groups



Compound #2

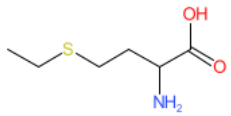
CAS: 1596-84-5
Dataset id:225 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : Carcinogen
Predicted value : Carcinogen

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



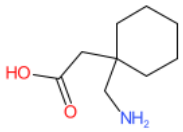
Compound #3

CAS: 628-94-4
Dataset id:782 (Training Set)
SMILES: O=C(N)CCCCC(=O)N
Similarity: 0.787
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



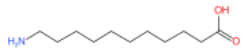
Compound #4

CAS: 13073-35-3
Dataset id:477 (Training Set)
SMILES: O=C(O)C(N)CCSCC
Similarity: 0.765
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 60142-96-3
Dataset id:789 (Training Set)
SMILES: O=C(O)CC1(CN)(CCCCC1)
Similarity: 0.757
Experimental value : Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 2432-99-7
Dataset id:36 (Training Set)
SMILES: O=C(O)CCCCCCCCCCN
Similarity: 0.75
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.799 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 = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

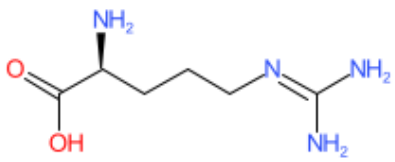
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 -

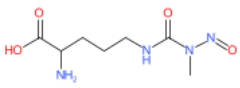


Prediction: Reliability:

Prediction is Possible NON-Carcinogen, but the result shows some critical aspects, which require to be checked:
- Only moderately similar compounds with known experimental value in the training set have been found
- similar molecules found in the training set have experimental values that disagree with the predicted value

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

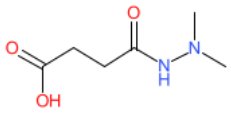
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 63642-17-1
Dataset id:590 (Training Set)
SMILES: O=NN(C(=O)NCCCC(N)C(=O)O)C
Similarity: 0.805
Experimental value : Carcinogen
Predicted value : Carcinogen

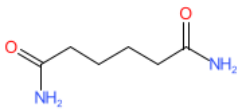
Alerts (not found also in the target): Carcinogenity alert no. 3; Carcinogenity alert no. 14; Carcinogenity alert no. 28



Compound #2

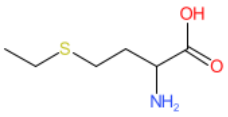
CAS: 1596-84-5
Dataset id:183 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : Carcinogen
Predicted value : Carcinogen

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



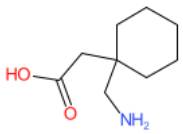
Compound #3

CAS: 628-94-4
Dataset id:602 (Training Set)
SMILES: O=C(N)CCCCC(=O)N
Similarity: 0.787
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #4

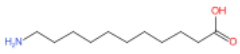
CAS: 67-21-0
Dataset id:14 (Training Set)
SMILES: O=C(O)C(N)CCSCC
Similarity: 0.765
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #5

CAS: 60142-96-3
Dataset id:725 (Training Set)
SMILES: O=C(O)CC1(CN)(CCCCC1)
Similarity: 0.757
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen

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



Compound #6

CAS: 2432-99-7
Dataset id:29 (Training Set)
SMILES: O=C(O)CCCCCCCCCCN
Similarity: 0.75
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.675 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.793 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 = 0.33 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✔	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

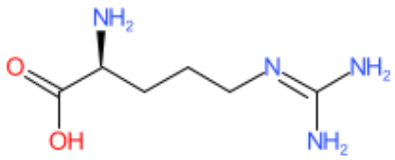
✔ The feature has a good assessment, model is reliable regarding this aspect.

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -

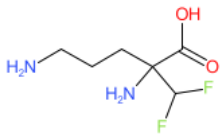


Prediction: Reliability:

Prediction is Carcinogen, but the result shows some critical aspects, which require to be checked:
- Accuracy of prediction for similar molecules found in the training set is not optimal
- similar molecules found in the training set have experimental values that disagree with the predicted value
The following relevant fragments have been found: Carcinogenity alert no. 127

Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Carcinogenic activity: Carcinogen
No. alerts for carcinogenicity: 1
Structural Alerts: Carcinogenity alert no. 127
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

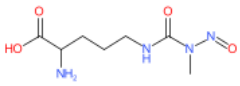
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 70052-12-9
Dataset id:245 (Training Set)
SMILES: O=C(O)C(N)(CCCN)C(F)F
Similarity: 0.85
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

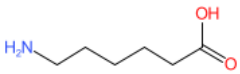
Alerts (not found also in the target): Carcinogenity alert no. 125



Compound #2

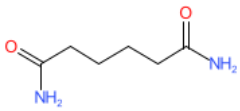
CAS: 63642-17-1
Dataset id:457 (Training Set)
SMILES: O=NN(C(=O)NCCCC(N)C(=O)O)C
Similarity: 0.805
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (not found also in the target): Carcinogenity alert no. 2; Carcinogenity alert no. 3; Carcinogenity alert no. 14; Carcinogenity alert no. 19; Carcinogenity alert no. 55; Carcinogenity alert no. 63



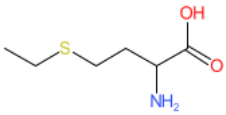
Compound #3

CAS: 60-32-2
Dataset id:47 (Training Set)
SMILES: O=C(O)CCCCCN
Similarity: 0.793
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #4

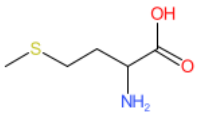
CAS: 628-94-4
Dataset id:23 (Training Set)
SMILES: O=C(N)CCCCC(=O)N
Similarity: 0.787
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen



Compound #5

CAS: 13073-35-3
Dataset id:299 (Test Set)
SMILES: O=C(O)C(N)CCSCC
Similarity: 0.765
Experimental value : Carcinogen
Predicted value : Possible NON-Carcinogen

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



Compound #6

CAS: 59-51-8
Dataset id:435 (Training Set)
SMILES: O=C(O)C(N)CCSC
Similarity: 0.765
Experimental value : NON-Carcinogen
Predicted value : Possible NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.613 Explanation: The predicted compound could be out of the Applicability Domain of the model.
✓	Similar molecules with known experimental value Similarity index = 0.813 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.649 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✗	Concordance for similar molecules Concordance index = 0.328 Explanation: similar molecules found in the training set have experimental values that disagree with the predicted value..
✓	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

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

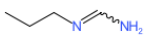
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

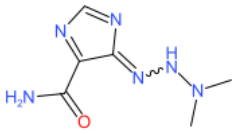
(Molecule 0) Reasoning on fragments/structural alerts :.

Fragment found: Carcinogenity alert no. 127



Structural alert for carcinogenity defined by the SMARTS: CCCN=CN

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



CAS: 4342-03-4
Dataset id:191 (Training Set)
SMILES: O=C(N)C1=NC=NC1(=NNN(C)C)
Similarity: 0.708

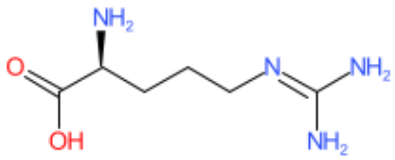
Experimental value : Carcinogen
Predicted value : Carcinogen

Alerts (found also in the target): Carcinogenity alert no. 127

Alerts (not found also in the target): Carcinogenity alert no. 55; Carcinogenity alert no. 79

1. Prediction Summary

Prediction for compound Molecule 0 -

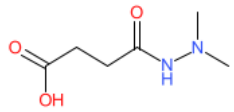


Prediction: Reliability:

Prediction is NON-Carcinogen, but the result shows some critical aspects, which require to be checked:
- Only moderately similar compounds with known experimental value in the training set have been found
- similar molecules found in the training set have experimental values that disagree with the predicted value

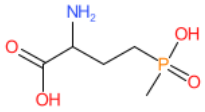
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Oral Carcinogenic class: NON-Carcinogen
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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



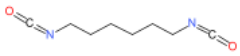
Compound #1

CAS: 1596-84-5
Dataset id:88 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : Carcinogen
Predicted value : Carcinogen



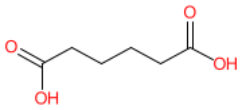
Compound #2

CAS: 77182-82-2
Dataset id:528 (Training Set)
SMILES: O=C(O)C(N)CCP(=O)(O)C
Similarity: 0.778
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



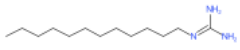
Compound #3

CAS: 822-06-0
Dataset id:538 (Training Set)
SMILES: O=C=NCCCCCCN=C=O
Similarity: 0.743
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



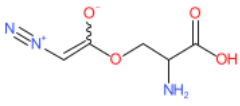
Compound #4

CAS: 124-04-9
Dataset id:541 (Training Set)
SMILES: O=C(O)CCCCC(=O)O
Similarity: 0.743
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 2439-10-3
Dataset id:490 (Training Set)
SMILES: N(=C(N)N)CCCCCCCCCCCC
Similarity: 0.741
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 115-02-6
Dataset id:27 (Training Set)
SMILES: N#[N+]C=C([O-])OCC(N)C(=O)O
Similarity: 0.74
Experimental value : Carcinogen
Predicted value : NON-Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.743 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.783 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 = 0.495 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 = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

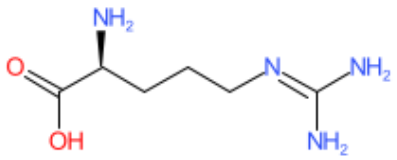
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 -

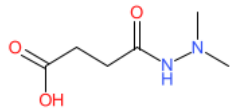


Prediction: Reliability:

Prediction is 0.39, 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
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)

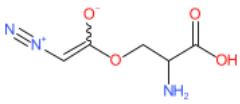
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Oral Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: 0.39
Predicted Oral Carcinogenicity SF [1/(mg/kg-day)]: 2.48
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



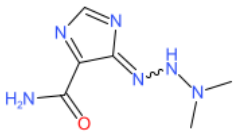
Compound #1

CAS: 1596-84-5
Dataset id:88 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : -1.74
Predicted value : -2.103



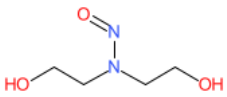
Compound #2

CAS: 115-02-6
Dataset id:27 (Training Set)
SMILES: N#[N+]C=C([O-])OCC(N)C(=O)O
Similarity: 0.74
Experimental value : 1.04
Predicted value : 0.917



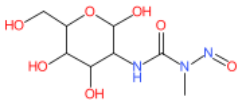
Compound #3

CAS: 4342-03-4
Dataset id:87 (Training Set)
SMILES: O=C(N)C1=NC=NC1(=NNN(C)C)
Similarity: 0.708
Experimental value : 1.69
Predicted value : 1.964



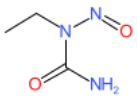
Compound #4

CAS: 1116-54-7
Dataset id:221 (Training Set)
SMILES: O=NN(CCO)CCO
Similarity: 0.68
Experimental value : 0.45
Predicted value : 0.521



Compound #5

CAS: 18883-66-4
Dataset id:278 (Test Set)
SMILES: O=NN(C(=O)NC1C(O)OC(CO)C(O)C1(O))C
Similarity: 0.672
Experimental value : 2.04
Predicted value : 2.315



Compound #6

CAS: 759-73-9
Dataset id:230 (Training Set)
SMILES: O=NN(C(=O)N)CC
Similarity: 0.664
Experimental value : 1.43
Predicted value : 1.378

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.389 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.763 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.243 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
⚠	Concordance for similar molecules Concordance index = 1.39 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 = 0.363 Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✖	Atom Centered Fragments similarity check ACF index = 0.51 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:

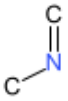
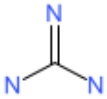
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

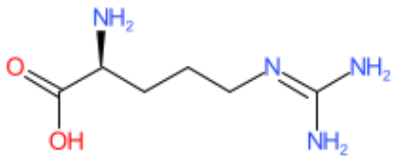
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

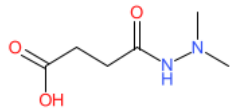


Prediction: Reliability:

Prediction is NON-Carcinogen, but the result shows some critical aspects, which require to be checked:
- 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

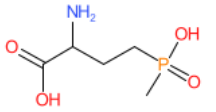
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Inhalation Carcinogenic class: NON-Carcinogen
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

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



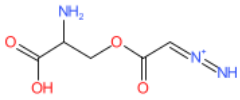
Compound #1

CAS: 1596-84-5
Dataset id:74 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : Carcinogen
Predicted value : Carcinogen



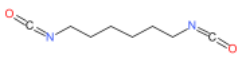
Compound #2

CAS: 77182-82-2
Dataset id:503 (Training Set)
SMILES: O=C(O)C(N)CCP(=O)(O)C
Similarity: 0.778
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



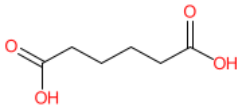
Compound #3

CAS: 115-02-6
Dataset id:24 (Training Set)
SMILES: O=C(O)C(N)COC(=O)C=[N+]=N
Similarity: 0.754
Experimental value : Carcinogen
Predicted value : Carcinogen



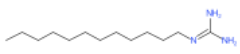
Compound #4

CAS: 822-06-0
Dataset id:514 (Training Set)
SMILES: O=C=NCCCCCCN=C=O
Similarity: 0.743
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #5

CAS: 124-04-9
Dataset id:517 (Training Set)
SMILES: O=C(O)CCCCC(=O)O
Similarity: 0.743
Experimental value : NON-Carcinogen
Predicted value : NON-Carcinogen



Compound #6

CAS: 2439-10-3
Dataset id:462 (Training Set)
SMILES: N(=C(N)N)CCCCCCCCCCCC
Similarity: 0.741
Experimental value : NON-Carcinogen
Predicted value : Carcinogen

3.2 Applicability Domain:
Measured Applicability Domain Scores

⚠	Global AD Index AD index = 0.743 Explanation: The predicted compound could be out of the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.783 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 = 0.495 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 = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

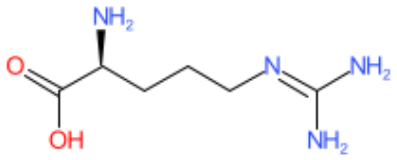
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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 -

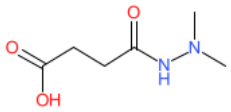


Prediction: Reliability:

Prediction is 0.79, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- Accuracy of prediction for similar molecules found in the training set is not optimal
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)

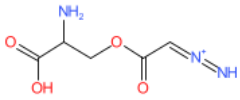
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted Inhalation Carcinogenicity SF (log form) [log(1/(mg/kg-day))]: 0.79
Predicted Inhalation Carcinogenicity SF [1/(mg/kg-day)]: 6.14
Experimental value [1/(mg/kg-day)]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



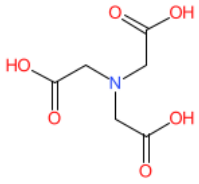
Compound #1

CAS: 1596-84-5
Dataset id:74 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : -1.75
Predicted value : -0.976



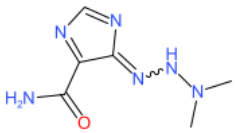
Compound #2

CAS: 115-02-6
Dataset id:24 (Training Set)
SMILES: O=C(O)C(N)COC(=O)C=[N+]=N
Similarity: 0.754
Experimental value : 1.04
Predicted value : -0.661



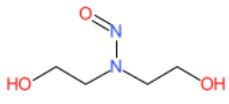
Compound #3

CAS: 139-13-9
Dataset id:178 (Training Set)
SMILES: O=C(O)CN(CC(=O)O)CC(=O)O
Similarity: 0.717
Experimental value : -2.28
Predicted value : 0.003



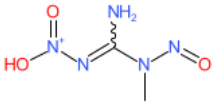
Compound #4

CAS: 4342-03-4
Dataset id:73 (Training Set)
SMILES: O=C(N)C1=NC=NC1(=NNN(C)C)
Similarity: 0.708
Experimental value : 1.69
Predicted value : 0.043



Compound #5

CAS: 1116-54-7
Dataset id:194 (Training Set)
SMILES: O=NN(CCO)CCO
Similarity: 0.68
Experimental value : 0.45
Predicted value : -0.147



Compound #6

CAS: 70-25-7
Dataset id:163 (Training Set)
SMILES: O=NN(C(=N[N+](=O)O)N)C
Similarity: 0.679
Experimental value : 0.92
Predicted value : 0.544

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.655 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.771 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 1.238 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
⚠	Concordance for similar molecules Concordance index = 1.395 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 = 1.701 Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
⚠	Atom Centered Fragments similarity check ACF index = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)..

Symbols explanation:

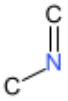
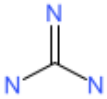
✔ The feature has a good assessment, model is reliable regarding this aspect.

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

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

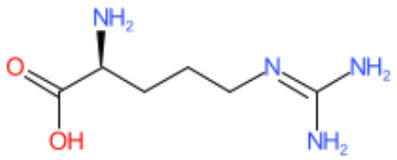
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

(Molecule 0) Reasoning on 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: CN=C The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: NC(N)=N The fragment has less than 3 occurrences in the model's training set

1. Prediction Summary

Prediction for compound Molecule 0 -

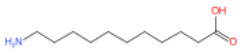


Prediction: Reliability:

Prediction is -2.7722, 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
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 unknown fragments and 1 infrequent_fragments found)

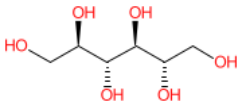
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: -2.7722
Predicted TD50 [mg/kg bw/day]: 591.85
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



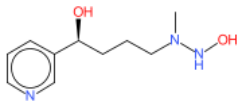
Compound #1

CAS: N.A.
Dataset id:34 (Training Set)
SMILES: C(CCCCCCCCCC(=O)O)N
Similarity: 0.75
Experimental value : -3.041
Predicted value : -2.357



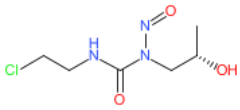
Compound #2

CAS: N.A.
Dataset id:162 (Test Set)
SMILES: C([C@@H]([C@@H]([C@@H]([C@@H](CO)O)O)O)O)O
Similarity: 0.735
Experimental value : -4.255
Predicted value : -3.919



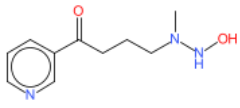
Compound #3

CAS: N.A.
Dataset id:68 (Training Set)
SMILES: c1ccc(cn1)[C@H](CCCN(NO)C)O
Similarity: 0.709
Experimental value : -0.86
Predicted value : 0.157



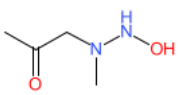
Compound #4

CAS: N.A.
Dataset id:155 (Test Set)
SMILES: N(C[C@@H](O)C)(C(=O)NCCCl)N=O
Similarity: 0.709
Experimental value : 0.065
Predicted value : -0.312



Compound #5

CAS: N.A.
Dataset id:52 (Training Set)
SMILES: c1ccc(cn1)C(=O)CCCN(NO)C
Similarity: 0.698
Experimental value : 0.186
Predicted value : -0.332



Compound #6

CAS: N.A.
Dataset id:2 (Training Set)
SMILES: CC(=O)CN(NO)C
Similarity: 0.686
Experimental value : 0.599
Predicted value : 0.545

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.379 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.742 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.51 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
⚠	Concordance for similar molecules Concordance index = 0.876 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 = 0.684 Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✖	Atom Centered Fragments similarity check ACF index = 0.51 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:

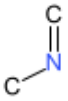
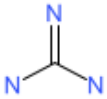
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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.

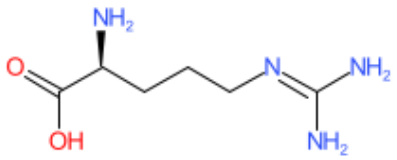
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

1. Prediction Summary

Prediction for compound Molecule 0 -

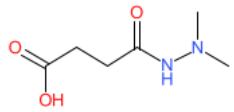


Prediction: Reliability:

Prediction is -4.4136, 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
- 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 (1 unknown fragments and 2 infrequent_fragments found)

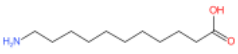
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted TD50 [-log(mg/kg bw/day)]: -4.4136
Predicted TD50 [mg/kg bw/day]: 25916.66
Experimental TD50 [mg/kg bw/day]: -
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



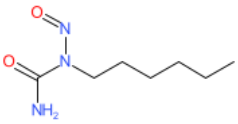
Compound #1

CAS: N.A.
Dataset id:111 (Training Set)
SMILES: C(=O)(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : -3.984
Predicted value : -2.932



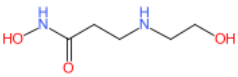
Compound #2

CAS: N.A.
Dataset id:117 (Training Set)
SMILES: C(CCCCCCCCCC(=O)O)N
Similarity: 0.75
Experimental value : -4.649
Predicted value : -4.782



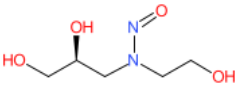
Compound #3

CAS: N.A.
Dataset id:154 (Test Set)
SMILES: C(CN(C(=O)N)N=O)CCCC
Similarity: 0.742
Experimental value : -1.124
Predicted value : -2.314



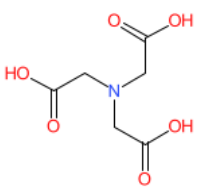
Compound #4

CAS: N.A.
Dataset id:80 (Training Set)
SMILES: N(CCO)CCC(=O)NO
Similarity: 0.735
Experimental value : -0.255
Predicted value : -2.037



Compound #5

CAS: N.A.
Dataset id:81 (Training Set)
SMILES: C([C@@H](CO)O)N(N=O)CCO
Similarity: 0.724
Experimental value : -0.777
Predicted value : -2.043



Compound #6

CAS: N.A.
Dataset id:98 (Training Set)
SMILES: N(CC(=O)O)(CC(=O)O)CC(=O)O
Similarity: 0.717
Experimental value : -2.908
Predicted value : -2.721

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.392 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.768 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.593 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✔	Concordance for similar molecules Concordance index = 0.333 Explanation: Similar molecules found in the training set have experimental values that agree with the predicted value..
⚠	Maximum error of prediction among similar molecules Max error index = 1.052 Explanation: the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
✖	Atom Centered Fragments similarity check ACF index = 0.51 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 2 infrequent_fragments found)..

Symbols explanation:

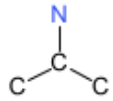
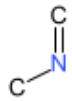
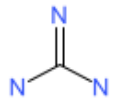
✔ The feature has a good assessment, model is reliable regarding this aspect.

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

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

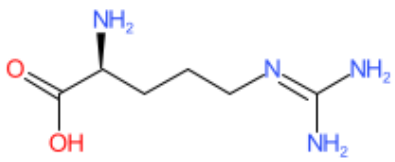
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

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

1. Prediction Summary

Prediction for compound Molecule 0 -

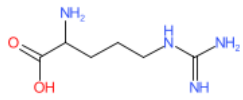


Prediction: Reliability:

Prediction is 9960.61 mg/kg, the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections.

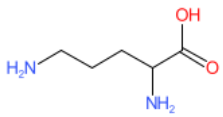
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted log LD50 [log(mmol/Kg)]: 1.757
Predicted log LD50 [mg/Kg]: 9960.61
Molecules used for prediction: 3
Experimental value [mg/Kg]: -
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

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



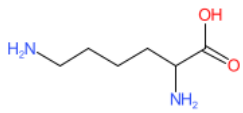
Compound #1

CAS: N.A.
Dataset id:3742 (Training Set)
SMILES: O=C(O)C(N)CCCNC(=N)N
Similarity: 0.906
Experimental value : 1.76
Predicted value : 1.755



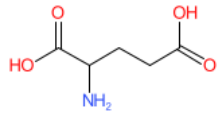
Compound #2

CAS: N.A.
Dataset id:5878 (Training Set)
SMILES: O=C(O)C(N)CCCN
Similarity: 0.865
Experimental value : 1.77
Predicted value : 1.317



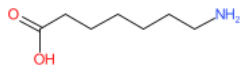
Compound #3

CAS: N.A.
Dataset id:5954 (Training Set)
SMILES: O=C(O)C(N)CCCCN
Similarity: 0.85
Experimental value : 1.74
Predicted value : 1.342



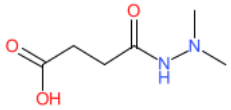
Compound #4

CAS: N.A.
Dataset id:3609 (Training Set)
SMILES: O=C(O)CCC(N)C(=O)O
Similarity: 0.809
Experimental value : 1.77
Predicted value : 1.061



Compound #5

CAS: N.A.
Dataset id:4078 (Training Set)
SMILES: O=C(O)CCCCCCN
Similarity: 0.79
Experimental value : 1.79
Predicted value : 1.603



Compound #6

CAS: N.A.
Dataset id:3686 (Training Set)
SMILES: O=C(O)CCC(=O)NN(C)C
Similarity: 0.789
Experimental value : 1.72
Predicted value : 1.501

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 = 0.871 Explanation: Strongly similar compounds with known experimental value in the training set have been ..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.285 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0.011 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 = 0.453 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:

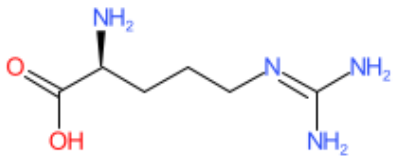
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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:

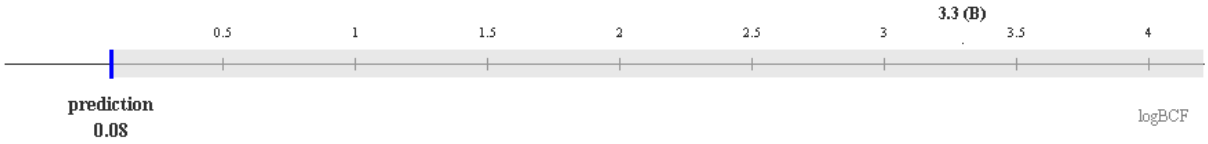
Prediction is 0.08 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- some atom centered 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: Carbonyl residue (SR 02); COOH group (PG 01)

Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted BCF [log(L/kg)]: 0.08
Predicted BCF [L/kg]: 1
Predicted BCF from sub-model 1 (HM) [log(L/kg)]: 0.19
Predicted BCF from sub-model 2 (GA) [log(L/kg)]: 0.23
Predicted LogP (MLogP): -0.72
Structural Alerts: Carbonyl residue (SR 02); COOH group (PG 01)
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

2. Possible Use and Uncertainty

Threshold 3.3 (bioaccumulative)

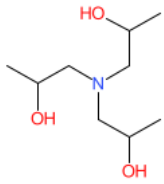
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.3, the current compound can not be associated (due to its Applicability Domain index value) to any conservative interval.
No safe classification can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



Threshold 3.7 (very bioaccumulative)

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

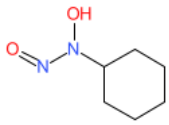
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 122-20-3
Dataset id:311 (Training Set)
SMILES: OC(C)CN(CC(O)C)CC(O)C
Similarity: 0.703
Experimental value : -0.24
Predicted value : 0.004

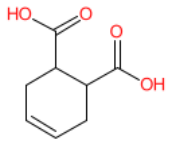
Alerts (not found also in the target): Tertiary amine (SR 05); OH group (PG 06)



Compound #2

CAS: 4883-72-1
Dataset id:434 (Training Set)
SMILES: O=NN(O)C1CCCCC1
Similarity: 0.694
Experimental value : 0.62
Predicted value : 0.304

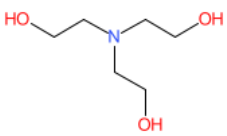
Alerts (not found also in the target): OH group (PG 06)



Compound #3

CAS: 88-98-2
Dataset id:121 (Training Set)
SMILES: O=C(O)C1CC=CCC1(C(=O)O)
Similarity: 0.687
Experimental value : 0.3
Predicted value : 0.107

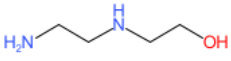
Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)



Compound #4

CAS: 102-71-6
Dataset id:310 (Test Set)
SMILES: OCCN(CCO)CCO
Similarity: 0.683
Experimental value : 0.59
Predicted value : 0.477

Alerts (not found also in the target): Tertiary amine (SR 05); OH group (PG 06)

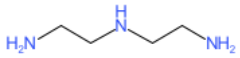


Compound #5

CAS: 111-41-1
Dataset id:309 (Training Set)
SMILES: OCCNCCN
Similarity: 0.678
Experimental value : 0.46
Predicted value : 0.502

Alerts (not found also in the target): OH group (PG 06); NH2 group (PG 07)

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



Compound #6

CAS: 111-40-0
Dataset id:307 (Training Set)
SMILES: NCCNCCN
Similarity: 0.666
Experimental value : 0.01
Predicted value : 0.404

Alerts (not found also in the target): NH2 group (PG 07)

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.594 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.698 Explanation: No similar compounds with known experimental value in the training set have been found..
✓	Accuracy of prediction for similar molecules Accuracy index = 0.28 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✓	Concordance for similar molecules Concordance index = 0.43 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 = 0.316 Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability..
✓	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set..
⚠	Atom Centered Fragments similarity check ACF index = 0.85 Explanation: some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)..

Symbols explanation:

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

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

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

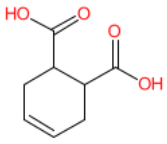
4.1 Reasoning:
Relevant Chemical Fragments and Moieties

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

Fragment found: Carbonyl residue (SR 02)

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

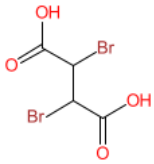
Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 88-98-2
Dataset id:121 (Training Set)
SMILES: O=C(O)C1CC=CCC1(C(=O)O)
Similarity: 0.687

Experimental value : 0.3
Predicted value : 0.107

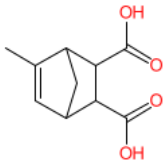
Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)



CAS: 526-78-3
Dataset id:71 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.66

Experimental value : 0.81
Predicted value : 0.227

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



CAS: 50853-70-8
Dataset id:131 (Training Set)
SMILES: O=C(O)C2C1C=C(C)C(C1)C2(C(=O)O)
Similarity: 0.652

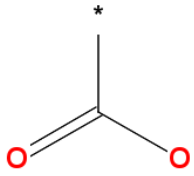
Experimental value : 0.74
Predicted value : 0.276

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

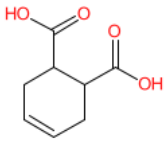
(Molecule 0) Reasoning on fragments/structural alerts - 2 of 2:.

Fragment found: COOH group (PG 01)



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

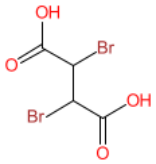
Following, the most similar compounds from the model's dataset having the same fragment.



CAS: 88-98-2
Dataset id:121 (Training Set)
SMILES: O=C(O)C1CC=CCC1(C(=O)O)
Similarity: 0.687

Experimental value : 0.3
Predicted value : 0.107

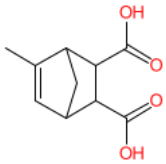
Alerts (found also in the target): Carbonyl residue (SR 02); COOH group (PG 01)



CAS: 526-78-3
Dataset id:71 (Training Set)
SMILES: O=C(O)C(C(C(=O)O)Br)Br
Similarity: 0.66

Experimental value : 0.81
Predicted value : 0.227

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



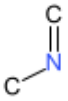
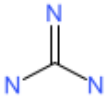
CAS: 50853-70-8
Dataset id:131 (Training Set)
SMILES: O=C(O)C2C1C=C(C)C(C1)C2(C(=O)O)
Similarity: 0.652

Experimental value : 0.74
Predicted value : 0.276

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

4.1 Reasoning:
Relevant Chemical Fragments and Moieties

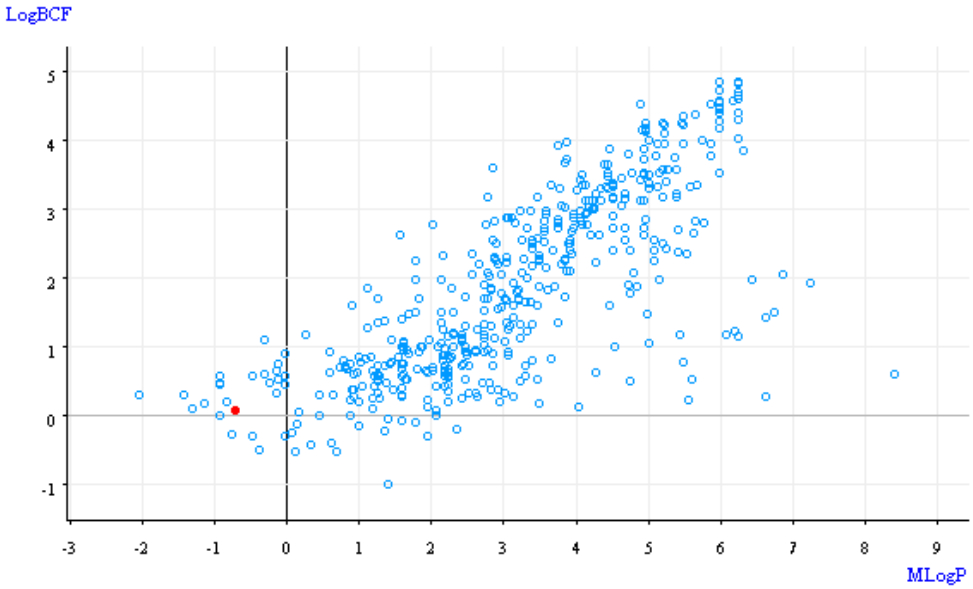
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CN=C The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: NC(N)=N The fragment has less than 3 occurrences in the model's training set

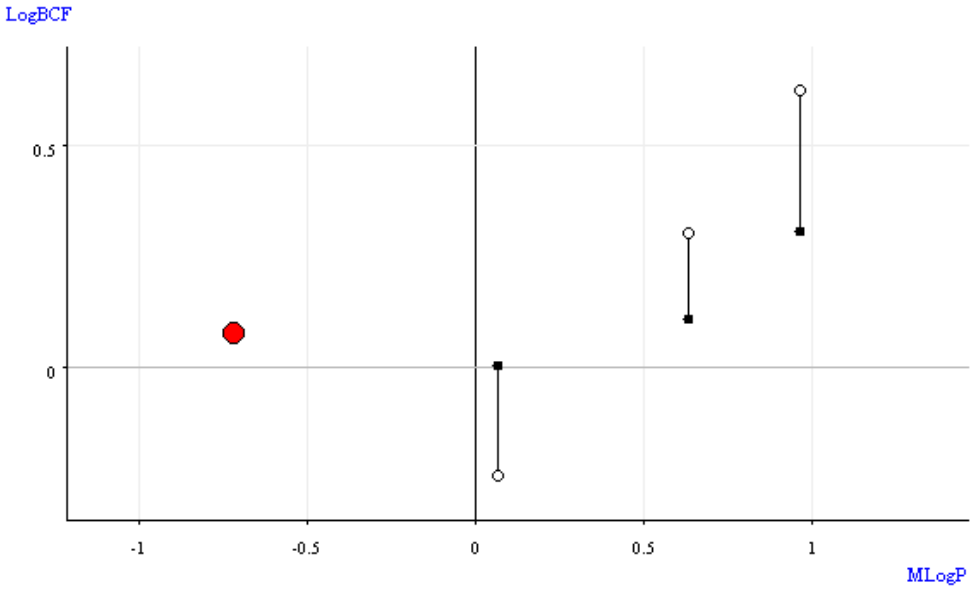
4.2 Reasoning:
Analysis of Molecular Descriptors

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

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

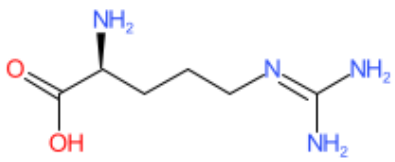


Following, a scatterplot of MLogP against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound;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



1. Prediction Summary

Prediction for compound Molecule 0 -

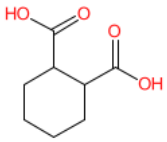


Prediction: Reliability:

Prediction is 0.5 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- No similar compounds with known experimental value in the training set have been found
- reliability of logP value used by the model is not optimal
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)

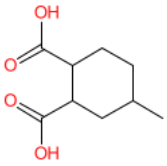
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted BCF [log(L/kg)]: 0.5
Predicted BCF [L/kg]: 3
Predicted LogP (Meylan/Kowwin): -3.92
Predicted LogP reliability: Moderate
MW: 173.68
Ionic compound: yes
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



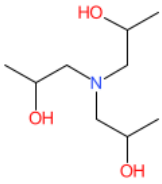
Compound #1

CAS: 1687-30-5
Dataset id:34 (Training Set)
SMILES: O=C(O)C1CCCCC1(C(=O)O)
Similarity: 0.718
Experimental value : 0.3
Predicted value : 0.5



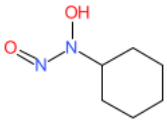
Compound #2

CAS: 57567-84-7
Dataset id:59 (Training Set)
SMILES: O=C(O)C1CCC(C)CC1(C(=O)O)
Similarity: 0.717
Experimental value : 0.38
Predicted value : 0.5



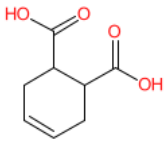
Compound #3

CAS: 122-20-3
Dataset id:627 (Test Set)
SMILES: OC(C)CN(CC(O)C)CC(O)C
Similarity: 0.703
Experimental value : -0.24
Predicted value : 0.5



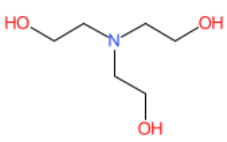
Compound #4

CAS: 4883-72-1
Dataset id:97 (Training Set)
SMILES: O=NN(O)C1CCCCC1
Similarity: 0.694
Experimental value : 0.6
Predicted value : 0.354



Compound #5

CAS: 88-98-2
Dataset id:12 (Training Set)
SMILES: O=C(O)C1CC=CCC1(C(=O)O)
Similarity: 0.687
Experimental value : 0.3
Predicted value : 0.5



Compound #6

CAS: 102-71-6
Dataset id:613 (Test Set)
SMILES: OCCN(CCO)CCO
Similarity: 0.683
Experimental value : 0.59
Predicted value : 0.5

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.61 Explanation: The predicted compound is outside the Applicability Domain of the model.
✖	Similar molecules with known experimental value Similarity index = 0.717 Explanation: No similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.16 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
✔	Concordance for similar molecules Concordance index = 0.16 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 = 0.2 Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability..
⚠	Reliability of logP prediction LogP reliability = 0.7 Explanation: reliability of logP value used by the model is not optimal..
✔	Model's descriptors range check Descriptors range check = True Explanation: descriptors for this compound have values inside the defined range..
⚠	Atom Centered Fragments similarity check ACF index = 0.85 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:

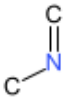
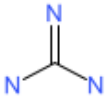
✔ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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

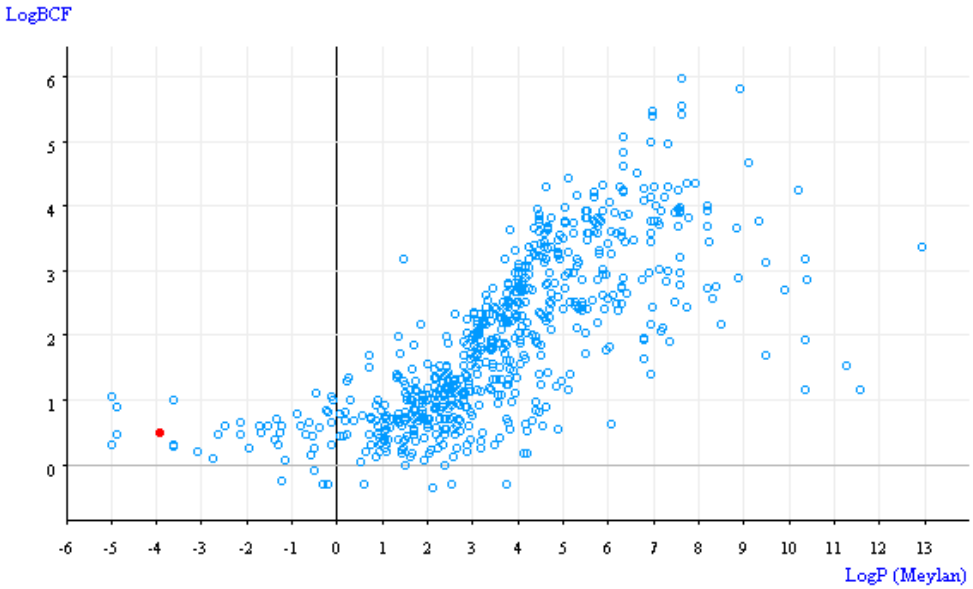
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CN=C The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: NC(N)=N The fragment has less than 3 occurrences in the model's training set

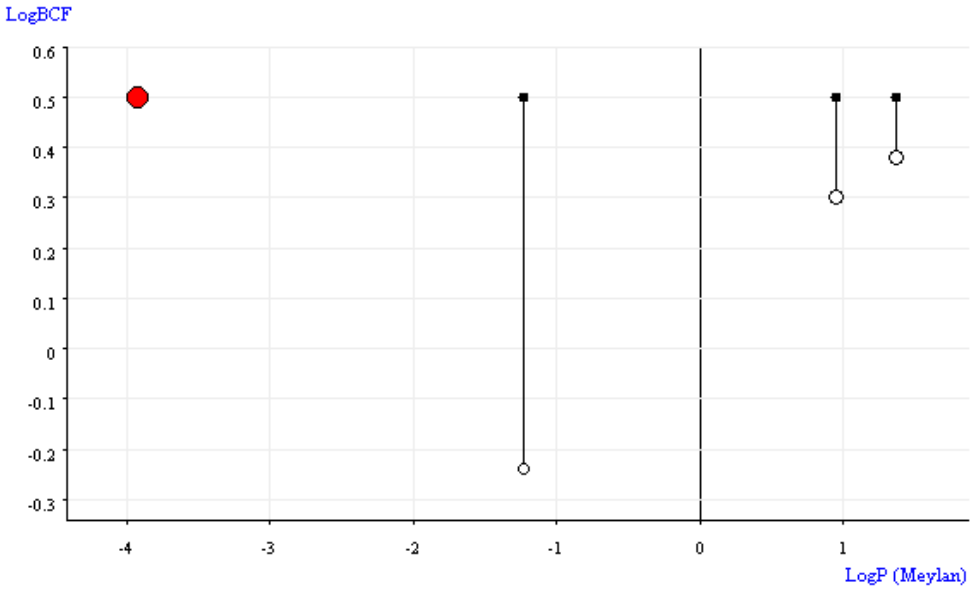
4.2 Reasoning:
Analysis of Molecular Descriptors

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

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

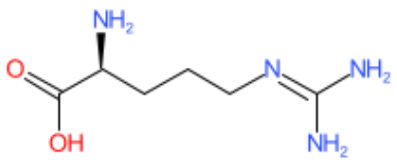


Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same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



1. Prediction Summary

Prediction for compound Molecule 0 -

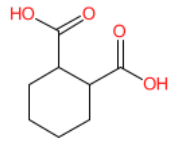


Prediction: Reliability:

Prediction is -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 optimal
- 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 optimal
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (2 infrequent_fragments found)

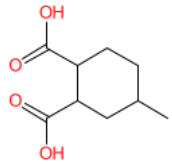
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted BCF (up) [log(L/kg)]: -0.05
Predicted BCF (up) [L/kg]: 0.89
Predicted BCF (low) [log(L/kg)]: -0.03
Predicted BCF (low) [L/kg]: 0.94
Predicted BCF (mid) [log(L/kg)]: -0.03
Predicted BCF (mid) [L/kg]: 0.93
Predicted LogP (Meylan/Kowwin): -3.92
Predicted LogP reliability: Moderate
Predicted kM (Meylan): -1.77
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



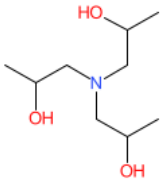
Compound #1

CAS: 1687-30-5
Dataset id:446 (Training Set)
SMILES: O=C(O)C1CCCCC1(C(=O)O)
Similarity: 0.718
Experimental value : -0.2
Predicted value : 0.242



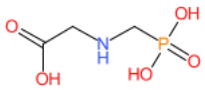
Compound #2

CAS: 57567-84-7
Dataset id:115 (Training Set)
SMILES: O=C(O)C1CCC(C)CC1(C(=O)O)
Similarity: 0.717
Experimental value : -0.16
Predicted value : 0.507



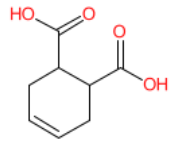
Compound #3

CAS: 122-20-3
Dataset id:751 (Training Set)
SMILES: OC(C)CN(CC(O)C)CC(O)C
Similarity: 0.703
Experimental value : -0.73
Predicted value : -0.049



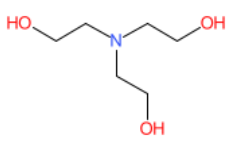
Compound #4

CAS: 1071-83-6
Dataset id:590 (Training Set)
SMILES: O=C(O)CNCP(=O)(O)O
Similarity: 0.702
Experimental value : 1.299
Predicted value : -0.049



Compound #5

CAS: 88-98-2
Dataset id:687 (Training Set)
SMILES: O=C(O)C1CC=CCC1(C(=O)O)
Similarity: 0.687
Experimental value : -0.2
Predicted value : 0.153



Compound #6

CAS: 102-71-6
Dataset id:362 (Training Set)
SMILES: OCCN(CCO)CCO
Similarity: 0.683
Experimental value : 0.1
Predicted value : -0.049

3.2 Applicability Domain:
Measured Applicability Domain Scores

✗	Global AD Index AD index = 0.61 Explanation: The predicted compound is outside the Applicability Domain of the model.
✗	Similar molecules with known experimental value Similarity index = 0.717 Explanation: No similar compounds with known experimental value in the training set have been found..
⚠	Accuracy of prediction for similar molecules Accuracy index = 0.555 Explanation: Accuracy of prediction for similar molecules found in the training set is not optimal..
✓	Concordance for similar molecules Concordance index = 0.131 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 = 0.667 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 = 0.7 Explanation: reliability of logP value used by the model is not optimal..
⚠	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:

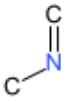
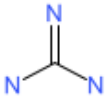
✓ The feature has a good assessment, model is reliable regarding this aspect.

⚠ 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

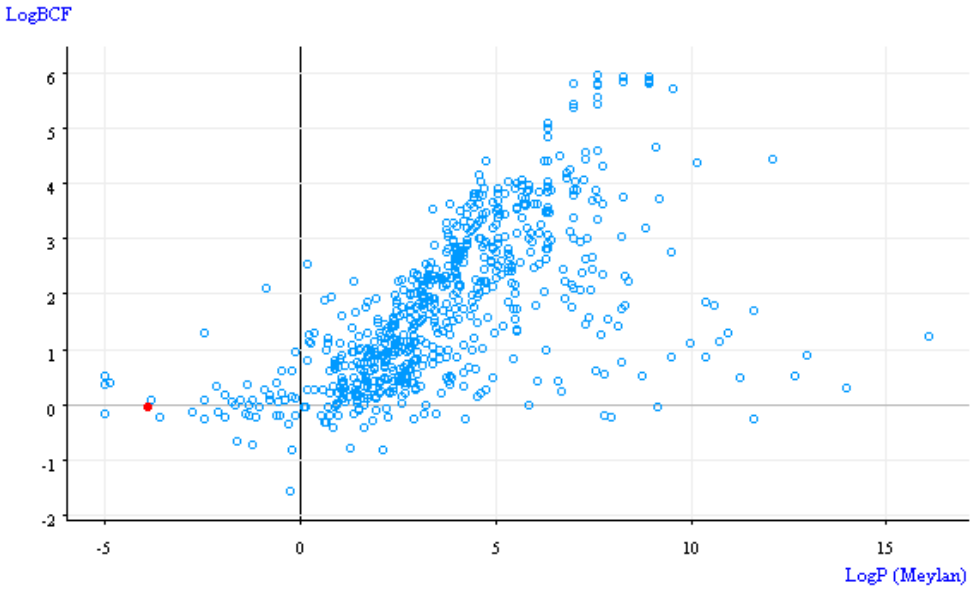
(Molecule 0) Reasoning on rare and missing Atom Centered Fragments .
The following Atom Centered Fragments have been found in the molecule, but they are not found or rarely found in the model's training set:

	Fragment defined by the SMILES: CN=C The fragment has less than 3 occurrences in the model's training set
	Fragment defined by the SMILES: NC(N)=N The fragment has less than 3 occurrences in the model's training set

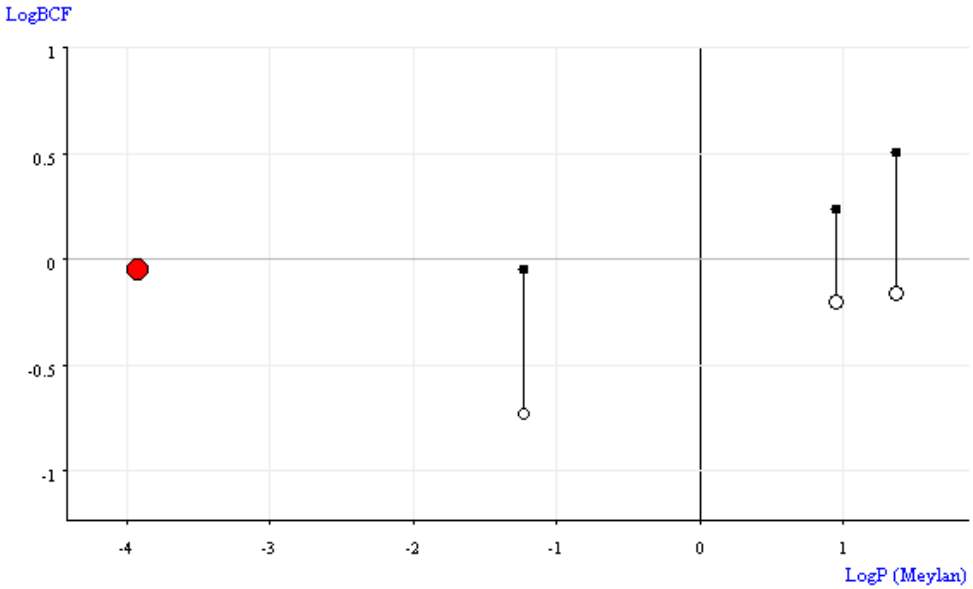
4.2 Reasoning:
Analysis of Molecular Descriptors

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

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

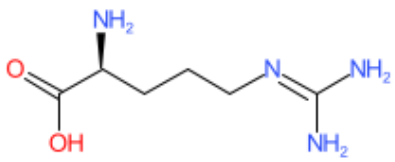


Following, a scatterplot of LogP (Meylan) against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same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



1. Prediction Summary

Prediction for compound Molecule 0 -

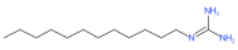


Prediction: Reliability:

Prediction is 0.07 log(L/kg), but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- Only moderately similar compounds with known experimental value in the training set have been found
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability

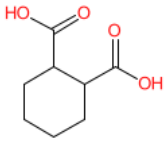
Compound: Molecule 0
Compound SMILES: O=C(O)C(N)CCCN=C(N)N
Experimental value: -
Predicted BCF [log(L/kg)]: 0.07
Molecules used for prediction: 4
Reliability: The predicted compound is outside the Applicability Domain of the model
Remarks:
none

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



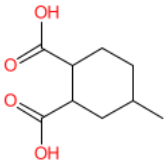
Compound #1

CAS: 2439-10-3
Dataset id:755 (Training Set)
SMILES: N(=C(N)N)CCCCCCCCCCCC
Similarity: 0.741
Experimental value : 1.204
Predicted value : 0.3



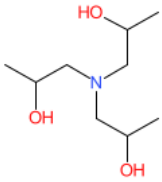
Compound #2

CAS: 610-09-3
Dataset id:354 (Training Set)
SMILES: O=C(O)C1CCCCC1(C(=O)O)
Similarity: 0.718
Experimental value : -0.199
Predicted value : -0.049



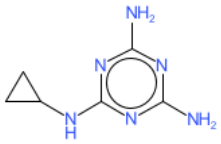
Compound #3

CAS: 57567-84-7
Dataset id:586 (Training Set)
SMILES: O=C(O)C1CCC(C)CC1(C(=O)O)
Similarity: 0.717
Experimental value : -0.159
Predicted value : -0.218



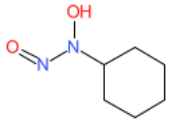
Compound #4

CAS: 122-20-3
Dataset id:257 (Training Set)
SMILES: OC(C)CN(CC(O)C)CC(O)C
Similarity: 0.703
Experimental value : -0.733
Predicted value : 0.544



Compound #5

CAS: 66215-27-8
Dataset id:654 (Training Set)
SMILES: n1c(nc(nc1N)NC2CC2)N
Similarity: 0.701
Experimental value : 0
Predicted value : 0.522



Compound #6

CAS: 4883-72-1
Dataset id:476 (Training Set)
SMILES: O=NN(O)C1CCCCC1
Similarity: 0.694
Experimental value : 0.291
Predicted value : 0.514

3.2 Applicability Domain:
Measured Applicability Domain Scores

✖	Global AD Index AD index = 0.7 Explanation: The predicted compound is outside the Applicability Domain of the model.
⚠	Similar molecules with known experimental value Similarity index = 0.719 Explanation: Only moderately similar compounds with known experimental value in the training set have been found..
✔	Accuracy of prediction for similar molecules Accuracy index = 0.598 Explanation: Accuracy of prediction for similar molecules found in the training set is good..
⚠	Concordance for similar molecules Concordance index = 0.608 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 = 1.277 Explanation: the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability..
✔	Atom Centered Fragments similarity check ACF index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Symbols explanation:

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

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

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

References and Documentation

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

Mutagenicity (Ames test) CONSENSUS model(version 1.0.4)

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

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

QSAR classification model for Mutagenicity (from CAESAR project)

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

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

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

QSAR classification model for Mutagenicity (SarPy/IRFMN)

References and Documentation

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

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

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

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

Carcinogenicity model (CAESAR)(version 2.1.10)

QSAR classification model for Carcinogenicity (from CAESAR project)

Carcinogenicity model (ISS)(version 1.0.3)

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

References and Documentation

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

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

Carcinogenicity model (IRFMN-Antares)(version 1.0.2)

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

Carcinogenicity oral classification model (IRFMN)(version 1.0.1)

Classification model for carcinogenicity (oral route).

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

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

References and Documentation

Carcinogenicity inhalation classification model (IRFMN)(version 1.0.1)

Classification model for carcinogenicity (inhalation route).

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

Quantitative model for the carcinogenicity inhalation route) Slope Factor.

Carcinogenicity in male rat (CORAL)(version 1.0.0)

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

Carcinogenicity in female Rat (CORAL)(version 1.0.0)

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

References and Documentation

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

KNN model for acute toxicity (LD50)

BCF model (CAESAR)(version 2.1.15)

QSAR regression model for fish BCF (from CAESAR project)

BCF model (Meylan)(version 1.0.4)

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

BCF model (Arnot-Gobas)(version 1.0.1)

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

References and Documentation

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

KNN (Read-Across) model for fish BCF