VEGA QSAR MODELS

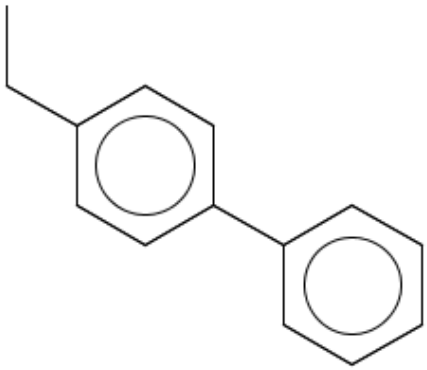
Report

Prediction and Applicability Domain analysis for models:

BCF model (CAESAR) (version 2.1.9)

Calculation core version: 1.0.26

(Mol. 1) Prediction for compound Molecule 1



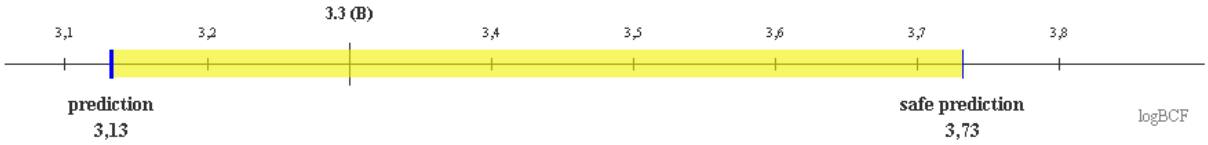
Model assessment: Prediction is logBCF = 3.13, the compound is IN the applicability domain and prediction seems to be reliable.

Compound: 1
Compound SMILES: c1ccc(cc1)c2ccc(cc2)CC
Experimental value: -
Prediction: 3.13 [log(L/kg)]
Prediction: 1356 [L/kg]
Prediction of model 1 (HM): 3.03 [log(L/kg)]
Prediction of model 2 (GA): 3.05 [log(L/kg)]
Structural Alerts: -
Calculated LogP: 4.41 [log units]
Reliability: Compound is in model Applicability Domain
Remarks for the prediction:
none

(Mol. 1) Class analysis of the compound:

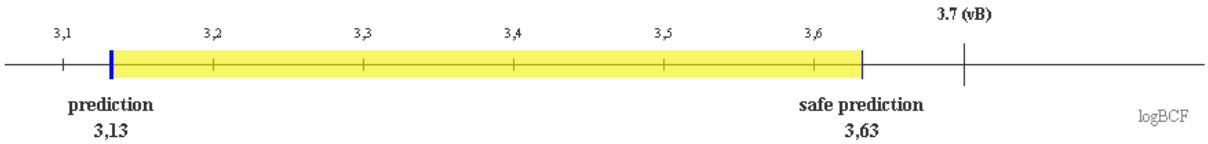
Threshold 3.3 (bioaccumulative)

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 be associated (due to its Applicability Domain index value) to a conservative interval of 0.6 log units.
On this basis, the compound can not be safely classified as not bioaccumulative.

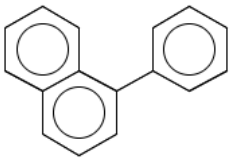


Threshold 3.7 (very bioaccumulative)

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 be associated (due to its Applicability Domain index value) to a conservative interval of 0.5 log units.
On this basis, the compound can not be safely classified as not very bioaccumulative.

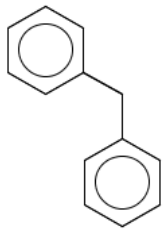


(Mol. 1) The following chemicals similar to the query compound have been found in the model's database:



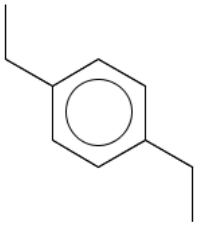
CAS: 605-02-7
Dataset id: 33 (training set)
SMILES: c1ccc(cc1)c3cccc2ccccc23
Similarity: 0.811

Experimental value: 3.51 [log(L/kg)]
Predicted value: 3.16 [log(L/kg)]



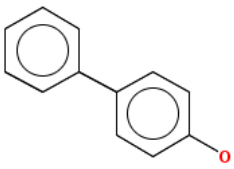
CAS: 101-81-5
Dataset id: 125 (training set)
SMILES: c1ccc(cc1)Cc2ccccc2
Similarity: 0.795

Experimental value: 2.94 [log(L/kg)]
Predicted value: 3.05 [log(L/kg)]



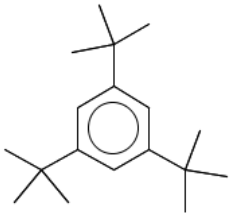
CAS: 105-05-5
Dataset id: 79 (training set)
SMILES: c1cc(ccc1CC)CC
Similarity: 0.795

Experimental value: 2.68 [log(L/kg)]
Predicted value: 2.64 [log(L/kg)]



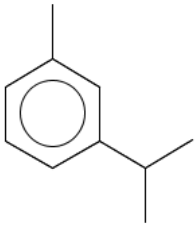
CAS: 92-69-3
Dataset id: 410 (test set)
SMILES: Oc1ccc(cc1)c2ccccc2
Similarity: 0.763

Experimental value: 1.59 [log(L/kg)]
Predicted value: 1.99 [log(L/kg)]



CAS: 1460-02-2
Dataset id: 291 (test set)
SMILES: c1c(cc(cc1C(C)(C)C)C(C)(C)C)C(C)(C)C
Similarity: 0.754

Experimental value: 4.37 [log(L/kg)]
Predicted value: 2.97 [log(L/kg)]



CAS: 535-77-3
Dataset id: 78 (training set)
SMILES: c1cc(cc(c1)C(C)C)C
Similarity: 0.749

Experimental value: 2.73 [log(L/kg)]
Predicted value: 2.63 [log(L/kg)]

(Mol. 1) Applicability Domain assessment details:

✓	**Global AD Index** AD Index = 1 Explanation: predicted substance is into the Applicability Domain of the model.
✓	**Similar molecules with known experimental value** Similarity index = 0.803 Explanation: strongly similar compounds with known experimental value in the training set have been found.
✓	**Accuracy (average error) of prediction for similar molecules** Accuracy index = 0.23 Explanation: accuracy of prediction for similar molecules found in the training set is good.
✓	**Concordance with similar molecules (average difference between target compound prediction and experimental values of similar molecules)** Concordance index = 0.285 Explanation: similar molecules found in the training set have experimental values that agree with the target compound predicted value.
✓	**Maximum error of prediction among similar molecules** Max error index = 0.35 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 matching index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.
✓	**Descriptors noise sensitivity analysis** Noise Sensitivity = 0.97 Explanation: predictions has a good response to noise scrambling, thus shows a good reliability.
✓	**Model 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.

Symbols explanation:

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

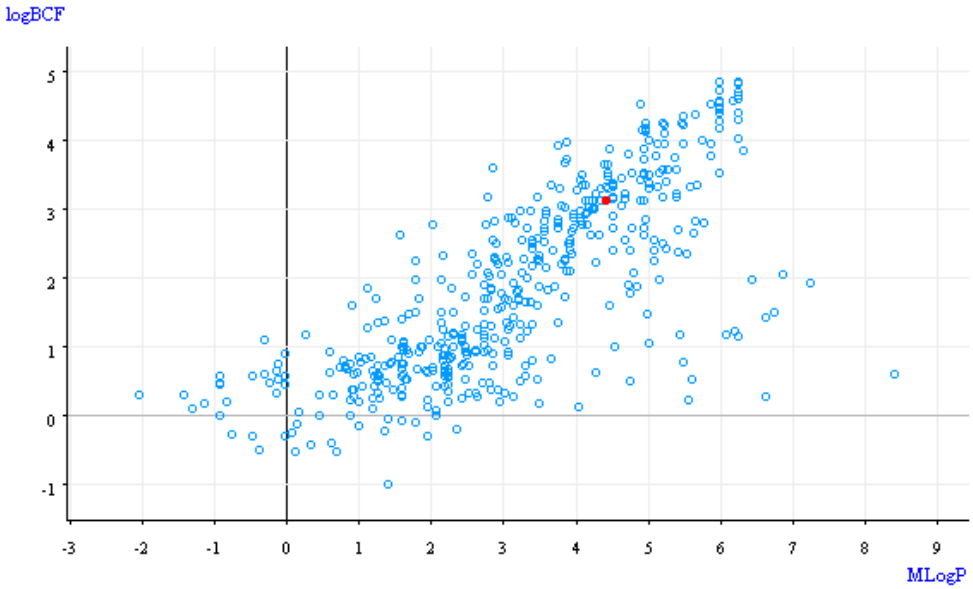
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

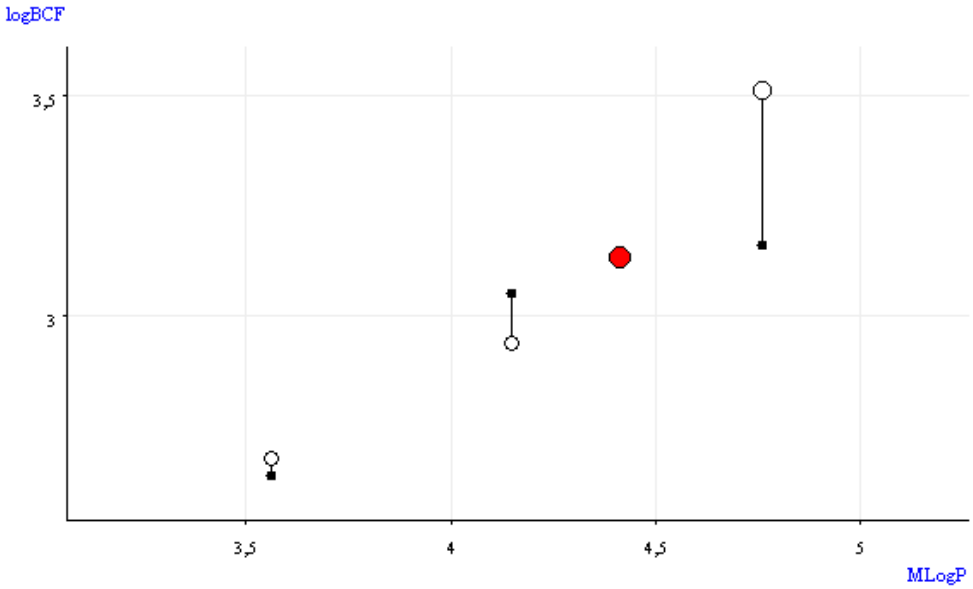
(Mol. 1) Reasoning on descriptors:

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

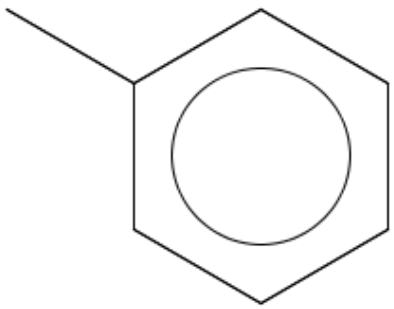
Following, a scatterplot of MLogP against response values; experimental values are reported for the training set, predicted value for the studied compound. Light blue dots represent values of compounds from training set, red dot is the value of the studied compound.



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.



(Mol. 2) Prediction for compound Molecule 2



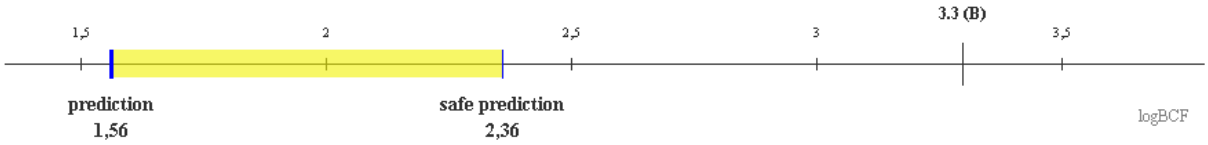
Model assessment: Prediction is logBCF = 1.56, but compound could be OUT of the applicability domain for the following reasons: - similar molecules found in the training set have experimental values that strongly disagree with the target compound predicted value

Compound: 2
Compound SMILES: c1ccc(cc1)C
Experimental value: -
Prediction: 1.56 [log(L/kg)]
Prediction: 36 [L/kg]
Prediction of model 1 (HM): 1.52 [log(L/kg)]
Prediction of model 2 (GA): 1.6 [log(L/kg)]
Structural Alerts: -
Calculated LogP: 2.61 [log units]
Reliability: Compound could be out of model Applicability Domain
Remarks for the prediction:
none

(Mol. 2) Class analysis of the compound:

Threshold 3.3 (bioaccumulative)

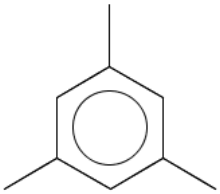
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.3, the current compound can be associated (due to its Applicability Domain index value) to a conservative interval of 0.8 log units.
On this basis, the compound can be safely classified as not bioaccumulative.



Threshold 3.7 (very bioaccumulative)

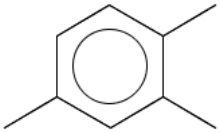
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.7, the current compound can be associated (due to its Applicability Domain index value) to a conservative interval of 0.9 log units.
On this basis, the compound can be safely classified as not very bioaccumulative.

(Mol. 2) The following chemicals similar to the query compound have been found in the model's database:



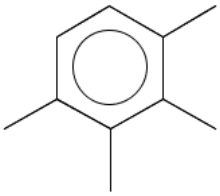
CAS: 108-67-8
Dataset id: 24 (training set)
SMILES: c1c(cc(cc1C)C)C
Similarity: 0.958

Experimental value: 2.55 [log(L/kg)]
Predicted value: 2.18 [log(L/kg)]



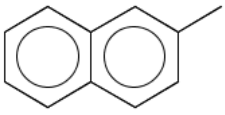
CAS: 95-63-6
Dataset id: 75 (training set)
SMILES: c1cc(c(cc1C)C)C
Similarity: 0.917

Experimental value: 2.08 [log(L/kg)]
Predicted value: 2.2 [log(L/kg)]



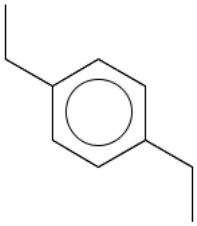
CAS: 488-23-3
Dataset id: 25 (training set)
SMILES: c1cc(c(c(c1C)C)C)C
Similarity: 0.908

Experimental value: 2.82 [log(L/kg)]
Predicted value: 2.57 [log(L/kg)]



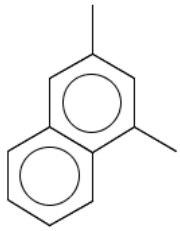
CAS: 91-57-6
Dataset id: 20 (test set)
SMILES: c1ccc2cc(ccc2(c1))C
Similarity: 0.848

Experimental value: 2.41 [log(L/kg)]
Predicted value: 2.72 [log(L/kg)]



CAS: 105-05-5
Dataset id: 79 (training set)
SMILES: c1cc(ccc1CC)CC
Similarity: 0.819

Experimental value: 2.68 [log(L/kg)]
Predicted value: 2.64 [log(L/kg)]



CAS: 575-41-7
Dataset id: 22 (training set)
SMILES: c1ccc2c(c1)cc(cc2C)C
Similarity: 0.818

Experimental value: 2.91 [log(L/kg)]
Predicted value: 2.91 [log(L/kg)]

(Mol. 2) Applicability Domain assessment details:

⚠	**Global AD Index** AD Index = 0.7 Explanation: predicted substance could be out of the Applicability Domain of the model.
✓	**Similar molecules with known experimental value** Similarity index = 0.936 Explanation: strongly similar compounds with known experimental value in the training set have been found.
✓	**Accuracy (average error) of prediction for similar molecules** Accuracy index = 0.245 Explanation: accuracy of prediction for similar molecules found in the training set is good.
✗	**Concordance with similar molecules (average difference between target compound prediction and experimental values of similar molecules)** Concordance index = 0.754 Explanation: similar molecules found in the training set have experimental values that strongly disagree with the target compound predicted value.
✓	**Maximum error of prediction among similar molecules** Max error index = 0.37 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 matching index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.
✓	**Descriptors noise sensitivity analysis** Noise Sensitivity = 0.899 Explanation: predictions has a good response to noise scrambling, thus shows a good reliability.
✓	**Model 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.

Symbols explanation:

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

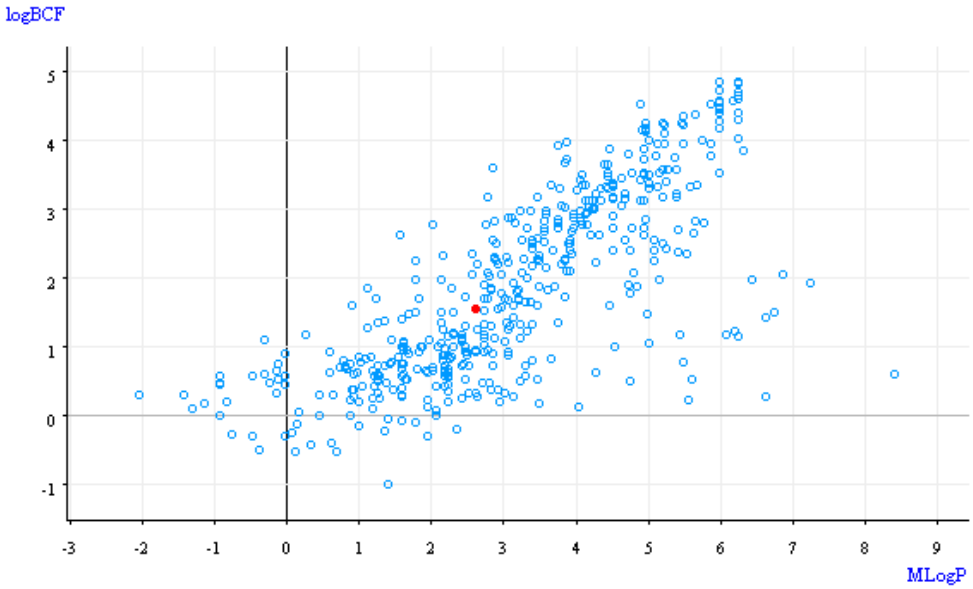
⚠ The feature has a non optimal assessment, this aspect should be reviewed by an expert.

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

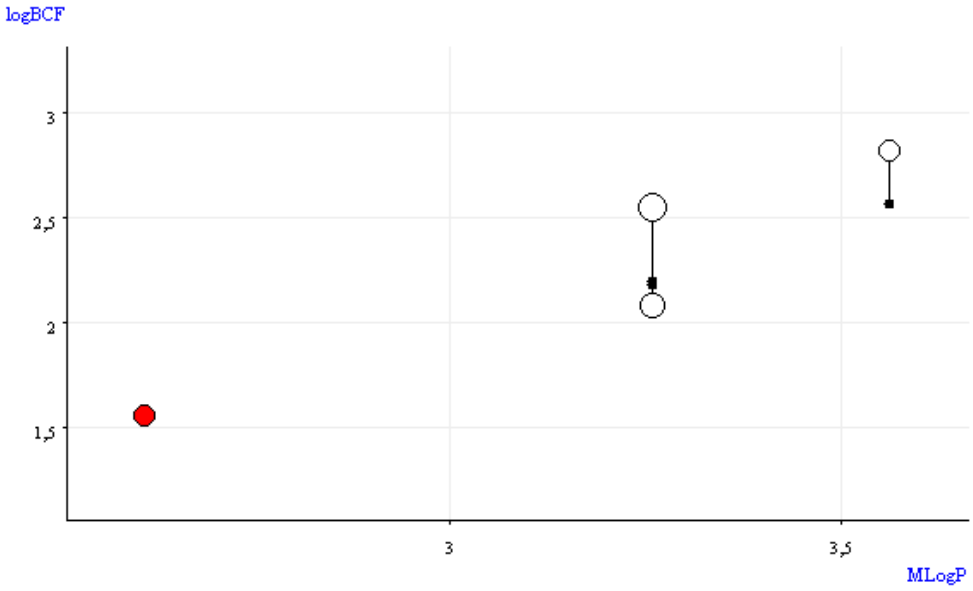
(Mol. 2) Reasoning on descriptors:

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

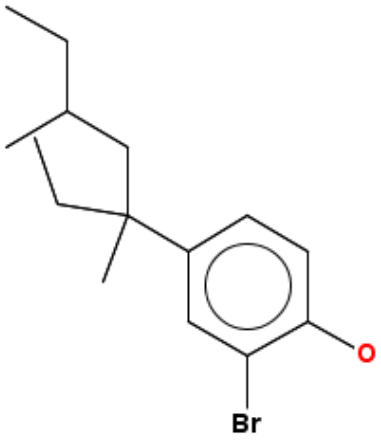
Following, a scatterplot of MLogP against response values; experimental values are reported for the training set, predicted value for the studied compound. Light blue dots represent values of compounds from training set, red dot is the value of the studied compound.



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.



(Mol. 3) Prediction for compound Molecule 3



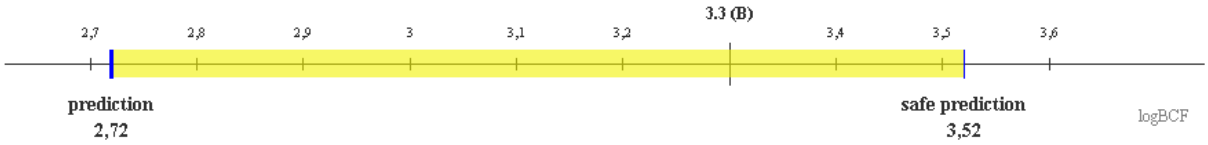
Model assessment: Prediction is logBCF = 2.72, but compound could be OUT of the applicability domain for the following reasons:
- accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that slightly disagree with the target compound predicted value
- the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
The following relevant fragments have been found: OH group (PG 06)

Compound: 3
Compound SMILES: Oc1ccc(cc1Br)C(C)(CC)CC(C)CC
Experimental value: -
Prediction: 2.72 [log(L/kg)]
Prediction: 524 [L/kg]
Prediction of model 1 (HM): 2.76 [log(L/kg)]
Prediction of model 2 (GA): 2.53 [log(L/kg)]
Structural Alerts: OH group (PG 06)
Calculated LogP: 4.82 [log units]
Reliability: Compound could be out of model Applicability Domain
Remarks for the prediction:
none

(Mol. 3) Class analysis of the compound:

Threshold 3.3 (bioaccumulative)

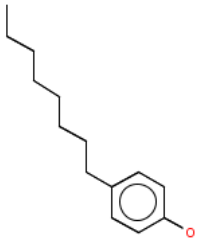
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.3, the current compound can be associated (due to its Applicability Domain index value) to a conservative interval of 0.8 log units.
On this basis, the compound can not be safely classified as not bioaccumulative.



Threshold 3.7 (very bioaccumulative)

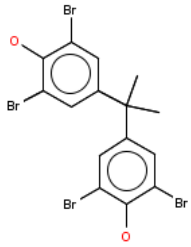
Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.7, the current compound can be associated (due to its Applicability Domain index value) to a conservative interval of 0.9 log units.
On this basis, the compound can not be safely classified as not very bioaccumulative.

(Mol. 3) The following chemicals similar to the query compound have been found in the model's database:



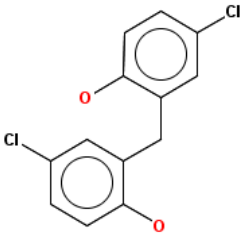
CAS: 1806-26-4
Dataset id: 85 (training set)
SMILES: Oc1ccc(cc1)CCCCCCCC
Similarity: 0.81

Experimental value: 1.87 [log(L/kg)]
Predicted value: 2.91 [log(L/kg)]



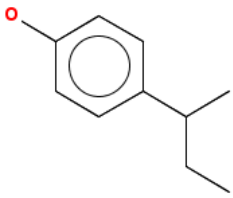
CAS: 79-94-7
Dataset id: 294 (training set)
SMILES: Oc1c(cc(cc1Br)C(c2cc(c(O)c(c2)Br)Br)(C)C)Br
Similarity: 0.807

Experimental value: 2.43 [log(L/kg)]
Predicted value: 1.74 [log(L/kg)]



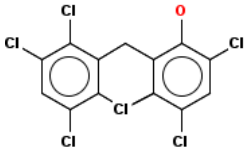
CAS: 97-23-4
Dataset id: 126 (training set)
SMILES: Oc1ccc(cc1Cc2cc(ccc2(O))Cl)Cl
Similarity: 0.755

Experimental value: 2.28 [log(L/kg)]
Predicted value: 2.37 [log(L/kg)]



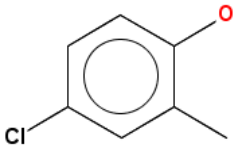
CAS: 99-71-8
Dataset id: 84 (training set)
SMILES: Oc1ccc(cc1)C(C)CC
Similarity: 0.745

Experimental value: 1.31 [log(L/kg)]
Predicted value: 1.47 [log(L/kg)]



CAS: 70-30-4
Dataset id: 127 (training set)
SMILES: Oc1c(cc(c(c1Cc2c(O)c(cc(c2Cl)Cl)Cl)Cl)Cl)Cl
Similarity: 0.734

Experimental value: 2.07 [log(L/kg)]
Predicted value: 3.14 [log(L/kg)]



CAS: 1570-64-5
Dataset id: 102 (training set)
SMILES: Oc1ccc(cc1C)Cl
Similarity: 0.729

Experimental value: 1.01 [log(L/kg)]
Predicted value: 0.98 [log(L/kg)]

(Mol. 3) Applicability Domain assessment details:

⚠	**Global AD Index** AD Index = 0.7 Explanation: predicted substance could be out of the Applicability Domain of the model.
✓	**Similar molecules with known experimental value** Similarity index = 0.808 Explanation: strongly similar compounds with known experimental value in the training set have been found.
✗	**Accuracy (average error) of prediction for similar molecules** Accuracy index = 0.865 Explanation: accuracy of prediction for similar molecules found in the training set is not adequate.
⚠	**Concordance with similar molecules (average difference between target compound prediction and experimental values of similar molecules)** Concordance index = 0.57 Explanation: similar molecules found in the training set have experimental values that slightly disagree with the target compound predicted value.
✗	**Maximum error of prediction among similar molecules** Max error index = 1.04 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 matching index = 1 Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.
✓	**Descriptors noise sensitivity analysis** Noise Sensitivity = 0.958 Explanation: predictions has a good response to noise scrambling, thus shows a good reliability.
✓	**Model 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.

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.

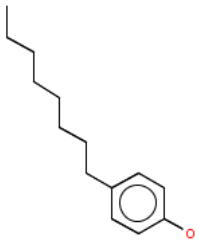
✗ The feature has a bad assessment, model is not reliable regarding this aspect.

(Mol. 3) Reasoning on fragments/structural alerts:

Fragment found: OH group (PG 06)

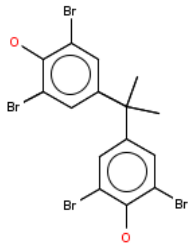
This chemical contains a OH polar group. The presence of polar groups increases hydrophilicity, related to lower values of BCF.

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



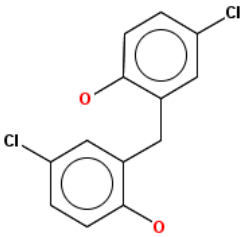
CAS: 1806-26-4
Dataset id: 85 (training set)
SMILES: Oc1ccc(cc1)CCCCCCCC
Similarity: 0.81

Experimental value: 1.87
Predicted value: 2.91



CAS: 79-94-7
Dataset id: 294 (training set)
SMILES: Oc1c(cc(cc1Br)C(c2cc(c(O)c(c2)Br)Br)(C)C)Br
Similarity: 0.807

Experimental value: 2.43
Predicted value: 1.74



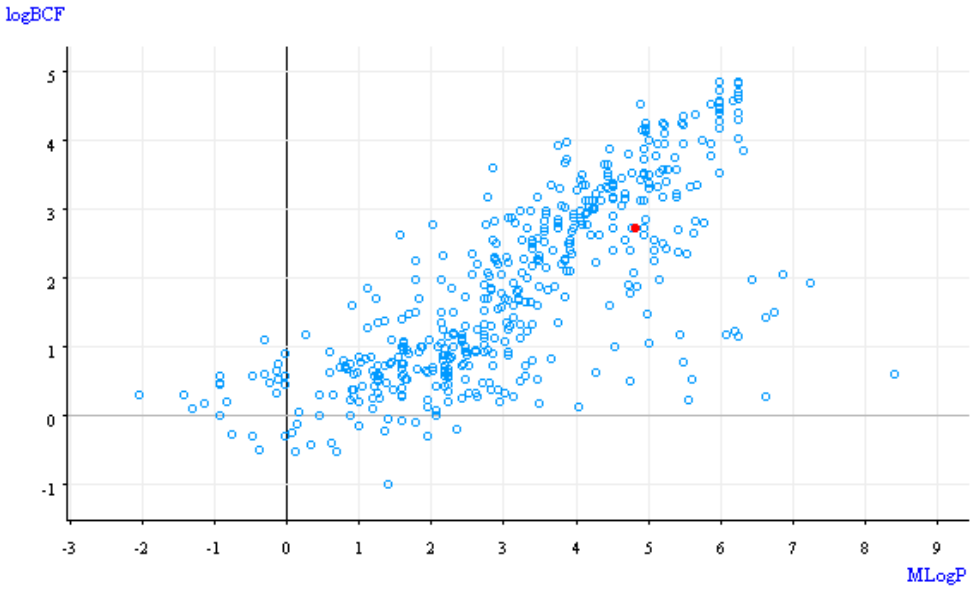
CAS: 97-23-4
Dataset id: 126 (training set)
SMILES: Oc1ccc(cc1Cc2cc(ccc2(O))Cl)Cl
Similarity: 0.755

Experimental value: 2.28
Predicted value: 2.37

(Mol. 3) Reasoning on descriptors:

Descriptor name: MLogP
Description: 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.