



[Center for Machine Learning and Intelligent Systems](#)

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

☒ Repository ☐ Web

[View ALL Data Sets](#)

Check out the beta version of the new UCI Machine Learning Repository we are currently testing! ✕
Contact us if you have any issues, questions, or concerns. [Click here to try out the new site.](#)

QSAR fish bioconcentration factor (BCF) Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Experimental bioconcentration factor (BCF) for 1056 molecules and binary fingerprints (extended connectivity) to be used for QSAR modeling.

Data Set Characteristics:	Multivariate	Number of Instances:	1056	Area:	Life
Attribute Characteristics:	Integer, Real	Number of Attributes:	7	Date Donated	2019-11-27
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	17710

Source:

Francesca Grisoni, University of Milano-Bicocca, Dept. of Earth and Environmental Sciences, Milano Chemometrics & QSAR Research Group, francesca.grisoni@unimib.it
Viviana Consonni, University of Milano-Bicocca, Dept. of Earth and Environmental Sciences, Milano Chemometrics & QSAR Research Group, viviana.consonni@unimib.it
Marco Vighi, University of Milano-Bicocca, Dept. of Earth and Environmental Sciences
Sara Villa, University of Milano-Bicocca, Dept. of Earth and Environmental Sciences
Roberto Todeschini, University of Milano-Bicocca, Dept. of Earth and Environmental Sciences, Milano Chemometrics & QSAR Research Group, roberto.todeschini@unimib.it

Data Set Information:

This dataset contains manually-curated experimental bioconcentration factor (BCF) for 1058 molecules (continuous values). Each row contains a molecule, identified by a CAS number, a name (if available), and a SMILES string. Additionally, the KOW (experimental or predicted) is reported. In this database, you will also find Extended Connectivity Fingerprints (binary vectors of 1024 bits), to be used as independent variables to predict the BCF. You can find additional information in the referenced papers.

In case you had questions, please do not hesitate to contact us!

Attribute Information:

The provided zip file contains two files.

(I) The file 'QSAR BCF KOW' contains the following attributes:

1. CAS number (molecule identifier)
2. Molecule Name (if not available, marked as 'n.a.')

3. SMILES string to identify the 2D molecular structure
4. LogKOW: octanol water partitioning coefficient (experimental or predicted, as indicated by the column 'KOW Type')
5. KOW Type: indicates whether the logKOW value is experimental or predicted
6. Experimental logBCF (quantitative response): experimental fish bioconcentration factor (logarithm form)

(II) The file 'ECFP_1024_m0-2_b2_c.txt' contains the following molecular descriptors (to be used to predict the BCF):
 - Extended Connectivity Fingerprints (ECFPs): binary descriptors useful to predict the experimental logBCF (computed with Dragon7, default settings --> details specified in the file)

Each row corresponds to one molecule, as identified by the SMILES field. The molecules are in the same order as in the previous file.

Relevant Papers:

1. Grisoni, F., Consonni, V., Villa, S., Vighi, M. and Todeschini, R., 2015. QSAR models for bioconcentration: Is the increase in the complexity justified by more accurate predictions?. Chemosphere, 127, pp.171-179. --> Procedure for data curation.
2. Grisoni, F., Consonni, V., Vighi, M., Villa, S. and Todeschini, R., 2016. Expert QSAR system for predicting the bioconcentration factor under the REACH regulation. Environmental research, 148, pp.507-512. --> Benchmark on the performance for this dataset
3. Grisoni, F., Consonni, V., Vighi, M., Villa, S. and Todeschini, R., 2016. Investigating the mechanisms of bioconcentration through QSAR classification trees. Environment international, 88, pp.198-205. --> Relationship between KOW and BCF

Citation Request:

If you publish results based on this dataset or parts of it, please cite the following paper:

```
@article{grisoni2015,
title={QSAR models for bioconcentration: Is the increase in the complexity justified by more accurate predictions?},
author={Grisoni, Francesca and Consonni, Viviana and Villa, Sara and Vighi, Marco and Todeschini, Roberto},
journal={Chemosphere},
volume={127},
pages={171--179},
year={2015},
publisher={Elsevier}
}
```

If you use the ECFP values, additionally please cite the following software:

Dragon (Software for Molecular Descriptor Calculation) Version 6.0 â€” 2012
[\[Web Link\]](#) (2012)

And paper:

```
@article{rogers2010,
title={Extended-connectivity fingerprints},
author={Rogers, David and Hahn, Mathew},
journal={Journal of chemical information and modeling},
volume={50},
number={5},
pages={742--754},
year={2010},
publisher={ACS Publications}
}
```

--> Thanks and happy predicting!

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)