



|   |  |   |
|---|--|---|
|  | <b>QMRF identifier (ECB Inventory):</b>  |  |
|   | <b>QMRF Title:</b> QSAR for bioconcentration (flow-through fish test) of polychlorinated biphenyls |   |
|   | <b>Printing Date:</b> Feb 16, 2010   |   |

## 1. QSAR identifier

### 1.1. QSAR identifier (title):

QSAR for bioconcentration (flow-through fish test) of polychlorinated biphenyls

### 1.2. Other related models:

### 1.3. Software coding the model:

QSARModel 3.3.5 Molcode Ltd., Turu 2, Tartu, 51014, Estonia <http://www.molcode.com>

## 2. General information

### 2.1. Date of QMRF:

05.06.2009

### 2.2. QMRF author(s) and contact details:

[1] Indrek Tulp Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[2] Tarmo Tamm Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[3] Gunnar Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[4] Dimitar Dobchev Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[5] Dana Martin Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[6] Kaido Tämm Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[7] Deniss Savchenko Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[8] Jaak Jänes Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[9] Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[10] Andres Kreegipuu Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

[11] Mati Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com)  
<http://www.molcode.com>

### 2.3. Date of QMRF update(s):

### 2.4. QMRF update(s):

### 2.5. Model developer(s) and contact details:

Molcode model development team Molcode Ltd. Turu 2, Tartu, 51014, Estonia [models@molcode.com](mailto:models@molcode.com) <http://www.molcode.com>

**2.6.Date of model development and/or publication:**

15.10.2008

**2.7.Reference(s) to main scientific papers and/or software package:**

[1]Karelson M, Dobchev D, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Karelson G (2008). Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors. ARKIVOC 16, 38-60.

[2]Karelson M, Karelson G, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Dobchev D (2009). QSAR study of pharmacological permeabilities. ARKIVOC 2, 218–238.

**2.8.Availability of information about the model:**

Model is proprietary, but the training and test sets are available.

**2.9.Availability of another QMRF for exactly the same model:**

None to date.

**3.Defining the endpoint - OECD Principle 1****3.1.Species:**

Various fish species (guppies, fathead minnow, rainbow trout, bluegill sunfish, etc)

**3.2.Endpoint:**

2.Environmental fate parameters 4.Bioconcentration 2.4.a.BCF fish

**3.3.Comment on endpoint:**

C.13 Bioconcentration, Fish BCF

**3.4.Endpoint units:****3.5.Dependent variable:**

$\log(\text{BCF})$  logarithm of the bioconcentration factor (BCF). The BCF for a particular chemical compound is defined as the equilibrium ratio of the concentration of a chemical inside an organism to the concentration in the surrounding environment.

**3.6.Experimental protocol:**

C.13 Bioconcentration: flow-through fish test Bioconcentration: flow-through fish test was determined using the OECD Test Guideline 305 (EU C.13). Bioconcentration is the process of accumulation of chemicals by fish through non-dietary routes under flow-through conditions and is one of the key steps through which chemicals are able to enter into the biosphere from physical surroundings. Bioconcentration factor (BCF) is the concentration of a particular chemical in a tissue per concentration of chemical in water ( $\text{BCF} = C_t/C_w$ ), being dimensionless. Due to wide use of chlorinated aromatic compounds in industry and the household (as bactericides, insecticides, herbicides, fungicides, and wood preservatives), they are an important source of soil and water contamination. The BCF data set, for 58 polychlorinated biphenyls (PCBs), was aggregated from various literature reports. The experimental  $\log\text{BCF}$  values for several fish species (guppies, fathead minnow, rainbow trout, and bluegill sunfish) were consolidated because there are too few data for each individual species. Furthermore, the PCB accumulation and metabolism mechanisms are assumed to be in all these species are similar, and the bioaccumulation factors depend mainly on the PCBs molecular structure and more particularly on the placement of each PCB in the biphenylchlorination reaction network described in the next subsections. Five sites in the Saginaw River were sampled in the summer of 1993: (1) the mouth of the Shiawassee River, (2) the mouth of the Tittabawassee River, (3) near the Zilwaukee Bridge, (4) at Buoy 26 upstream of Middleground Island (site of dredge spoil disposal), (5) downstream of Bay City,

near Buoy 10A. Three techniques of assessing bioavailable polychlorinated biphenyls (PCBs) in the Saginaw River, MI, were compared: sediments, caged fish, and semipermeable membrane devices (SPMDs). SPMDs (2.5 cm x 152 cm, 85- $\mu$ m membrane thickness, total mass 8.35 g each) were assembled from low-density polyethylene lay-flat tubing containing a thin film of 95% pure triolein (1.64 g), the ends of which were heat-sealed. Each field sample consisted of four SPMDs, and each field blank sample consisted of two SPMDs. The SPMDs were sealed in new, hexane-rinsed, metal cans and were stored at -20 °C prior to field deployment. The SPMDs were deployed in stainless steel cages. Field blanks, which accompanied the field samplers to the sampling sites, were removed during the deployment interval, then replaced in their containers and returned to the laboratory. The field blanks were used to determine any background contamination as SPMDs were deployed and retrieved. The exposure period for the field SPMDs was 28 days at mid-depth in the water column, ranging 1-5 m above the sediment. Twenty juvenile, hatchery-reared, channel catfish (*Ictalurus punctatus*, ~ 8-10 cm length) were placed in cages at each site for 28 days. These cages were within a few meters of the SPMD cages at the same depth. A subset of the catfish were used as reference samples at the beginning of the study. The samples were analyzed for PCB congeners to determine concentrations and patterns. Total PCB concentrations ranged from 33 to 280 ng/g (dry weight) in sediments, 46 to 290 ng/g (wet weight) in caged fish, and 77 to 790 ng/g in SPMDs. References 1-9 in section 9.2 (bibliography).

### **3.7.Endpoint data quality and variability:**

Experimental data were taken Qin et al (2008). Statistics: max value: 5.92 min value: 2.64 standard deviation: 0.895 skewness: -0.810

## **4.Defining the algorithm - OECD Principle 2**

### **4.1.Type of model:**

QSAR

### **4.2.Explicit algorithm:**

Multilinear regression QSAR

$\log BCF = 6.598 * \text{Average Bonding Information content (order 1)} - 291.721 * \text{Partial Charged (Zefirov) Surface Area of H atoms} - 9.189 * \text{Polarity parameter (Zefirov) / distance} + 12.497$

### **4.3.Descriptors in the model:**

[1]Average Bonding Information content (order 1) (the average information content divided by  $\log 2q$ , where  $q$  is the number of edges in the structural graph of a molecule)

[2]Partial Charged (Zefirov) Surface Area of H atoms (calculated as a combination of the contribution of charged H atom areas to the total molecular solvent accessible surface area)

[3]Polarity parameter (Zefirov) / distance (polarity parameter factorized by the division with the distance between atoms bearing minimum and maximum partial charges) [ $1/\text{\AA}$ ]

### **4.4.Descriptor selection:**

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules: 1-parameter equations: Fisher criterion and  $R^2$  over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold; 2-parameter equations: intercorrelation coefficient below threshold, significant correlation with endpoint in terms of correlation coefficient and t-test. Stepwise trial of additional descriptors not significantly correlated to any already in the model.

#### **4.5. Algorithm and descriptor generation:**

1D, 2D, and 3D theoretical calculations quantum chemical descriptors derived from MMFFs (vacuum) conformational search and AM1 calculation. Model developed by using multilinear regression.

#### **4.6. Software name and version for descriptor generation:**

QSARModel 3.3.5

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

<http://www.molcode.com>

#### **4.7. Descriptors/Chemicals ratio:**

19.33 (58 chemicals / 3 descriptors)

### **5. Defining the applicability domain - OECD Principle 3**

#### **5.1. Description of the applicability domain of the model:**

Applicability domain based on training set. By chemical identity: polychlorinated biphenyls (PCBs), with no other functional groups. By descriptor value range: the model is suitable for compounds that have the descriptors in the following range: Average Bonding Information content (order 1) (min: 0.21, max: 0.65), Partial Charged (Zefirov) Surface Area of H atoms (min: 0.001, max: 0.016), Polarity parameter (Zefirov) / distance (min: 0.008, max: 0.103)

#### **5.2. Method used to assess the applicability domain:**

Presence of functional groups in structures. Range of descriptor values in training set with  $\pm 30\%$  confidence. Descriptor values must fall between maximal and minimal descriptor values of training set  $\pm 30\%$ .

#### **5.3. Software name and version for applicability domain assessment:**

QSARModel 3.3.5

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

<http://www.molcode.com>

#### **5.4. Limits of applicability:**

### **6. Internal validation - OECD Principle 4**

#### **6.1. Availability of the training set:**

Yes

#### **6.2. Available information for the training set:**

CAS RN: Yes

Chemical Name: Yes

Smiles: No

Formula: No

INChI: No

MOL file: Yes

#### **6.3. Data for each descriptor variable for the training set:**

All

**6.4.Data for the dependent variable for the training set:**

All

**6.5.Other information about the training set:**

52 data points: 0 negative values; 52 positive values

**6.6.Pre-processing of data before modelling:**

**6.7.Statistics for goodness-of-fit:**

$R^2 = 0.945$  (Correlation coefficient);  $s = 0.047$  (Standard error of the estimate);  $F = 272.996$  (Fisher function);

**6.8.Robustness - Statistics obtained by leave-one-out cross-validation:**

$R^2_{cv} = 0.933$  LOO;

**6.9.Robustness - Statistics obtained by leave-many-out cross-validation:**

$R^2_{cv} = 0.930$  LMO; 80%: 20%

**6.10.Robustness - Statistics obtained by Y-scrambling:**

**6.11.Robustness - Statistics obtained by bootstrap:**

**6.12.Robustness - Statistics obtained by other methods:**

ABC analysis (2:1 training : prediction) on sorted data divided into 3 subsets (A;B;C). Training set formed with 2/3 of the compounds (set A+B, A+C, B+C) and validation set consisted of 1/3 of the compounds (C, B, A) average  $R^2$  (fitting) = 0.943 average  $R^2$  (prediction) = 0.933

**7.External validation - OECD Principle 4**

**7.1.Availability of the external validation set:**

Yes

**7.2.Available information for the external validation set:**

CAS RN:Yes

Chemical Name:Yes

Smiles:No

Formula:No

INChI:No

MOL file:Yes

**7.3.Data for each descriptor variable for the external validation set:**

All

**7.4.Data for the dependent variable for the external validation set:**

All

**7.5.Other information about the external validation set:**

6 data points: 0 negative values; 6 positive values

**7.6.Experimental design of test set:**

The full experimental dataset was sorted according to increasing values of logBCF and each tenth compound was assigned to the test set.

**7.7.Predictivity - Statistics obtained by external validation:**

$R^2=0.959$

### 7.8. Predictivity - Assessment of the external validation set:

The descriptors for the test set are in the limits of the applicability domain.

### 7.9. Comments on the external validation of the model:

## 8. Providing a mechanistic interpretation - OECD Principle 5

### 8.1. Mechanistic basis of the model:

The BCF depends on the shape, and the partial charges and polarity of the compound and the way these charges are distributed on the surface of the molecule.

### 8.2. A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments.

### 8.3. Other information about the mechanistic interpretation:

The descriptor Average Bonding Information content (order 1) gives information about the spatial shape of the molecule (including branching and symmetry). The positive coefficient indicates that fully substituted (non-symmetric) molecules lead to lower BCF. The descriptors Partial Charged (Zefirov) Surface Area of H atoms and Polarity parameter (Zefirov) / distance reflect the charge distribution of the molecule. Both share negative coefficients, indicating that highly polar hydrogen atoms on the surface of the molecules and large separated partial charges have negative influence on BCF. The proposed mechanism based on the model agrees well with literature (Qin et al, 2008).

## 9. Miscellaneous information

### 9.1. Comments:

Polychlorinated biphenyls (PCBs) are some of the most prevalent pollutants in the total environment and receive more and more concerns as a group of ubiquitous potential persistent organic pollutants.

### 9.2. Bibliography:

- [1] Ivanciuc T, Ivanciuc O & Klein DJ (2006). Modeling the bioconcentration factors and bioaccumulation factors of polychlorinated biphenyls with posetic quantitative super-structure/activity relationships (QSSAR). *Molecular Diversity* 10, 133–145.
- [2] Echols KR, Gale RW, Schwartz TR, Huckins JN, Williams LL, Meadows JC, Morse D, Petty JD, Orazio CE & Tillitt DE (2000). Comparing polychlorinated biphenyl concentrations and patterns in the Saginaw River using sediment caged fish, and semipermeable membrane devices. *Environmental Science and Technology* 34, 4095–410.
- [3] Geyer HJ, Scheunert I, Bruggemann R, Steinberg C, Korte F & Kettrup A (1991). QSAR for organic chemical bioconcentration in *Daphnia*, algae, and mussels. *Science of the Total Environment* 109/110, 387–394.
- [4] Devillers J, Bintein S & Domine D (1996). Comparison of BCF models based on log P. *Chemosphere* 33, 1047–1065.
- [5] Wei D, Zhang A, Wu C, Han S & Wang L (2001). Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compound (PCOCs). *Chemosphere* 44, 1421–1428.
- [6] Saçan MT, Erdem SS, Özpınar GA & Balcioglu IA (2004). QSPR study on the bioconcentration factors of nonionic organic compounds in fish by characteristic root index and semiempirical molecular descriptors. *Journal of Chemical Information and Computer Sciences*

44, 985–992.

[7]Lu X, Tao S, Cao J & Dawson RW (1999). Prediction of fish bioconcentration factors of nonpolar organic pollutants based on molecular connectivity indices. *Chemosphere* 39, 987–999.

[8]Lu X, Tao S, Hu H & Dawson RW (2000). Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors. *Chemosphere* 41, 1675–1688.

[9]Fox K, Zauke GP and Butte W (1994). Kinetics of bioconcentration and clearance of 28 polychlorinated biphenyl congeners in Zebrafish (*Brachydanio rerio*). *Ecotoxicology and Environmental Safety* 28, 99–109.

[10]Qin LT, Liu SS, Liu HL & Ge HL (2008). A new predictive model for the bioconcentration factors of polychlorinated biphenyls (PCBs) based on the molecular electronegativity distance vector (MEDV). *Chemosphere* 70, 1577-1587.

### 9.3.Supporting information:

Training set(s)

|                                |   |
|--------------------------------|---|
| BCF polychlorinate training 58 | <a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf173_BCF_polychlorinate_training_58.sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf173_BCF_polychlorinate_training_58.sdf</a> |
|--------------------------------|---|

Test set(s)

|                           |   |
|---------------------------|---|
| BCF polychlorinate test 6 | <a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf173_BCF_polychlorinate_test_6.sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf173_BCF_polychlorinate_test_6.sdf</a> |
|---------------------------|---|

## 10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

BCF, bioconcentration, flow through test, fish, Molcode

10.4.Comments: