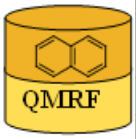	*QMRF identifier (JRC Inventory):* *QMRF Title:* *QSAR model for Fish, early-life stage toxicity test* *Printing Date:*9.12.2010	

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR model for Fish, early-life stage toxicity test

1.2.Other related models:

1.3.Software coding the model:

QSARModel 5.0.0 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

08.12.2010

2.2.QMRF author(s) and contact details:

[1]Indrek Tulp Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[2]Tarmo Tamm Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[3]Gunnar Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[4]Dimitar Dobchev Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[5]Kaido Tämm Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[6]Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia 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 http://www.molcode.com

2.6.Date of model development and/or publication:

03.12.2010

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:

All information in full detail is 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:

Zebra fish (Brachydanio rerio), fathead minnow (Pimephales promelas)

3.2.Endpoint:

3.Ecotoxic effects 3.5.Long-term toxicity to fish (egg/sac fry, growth inhibition of juvenile fish, early life stage, full life cycle)

3.3.Comment on endpoint:

Determination of early-life stage toxicity

Early-life stage toxicity was determined using the OECD test guideline 210. The early-life stages of fish are exposed to a range of concentrations of the test substance dissolved in water, preferably under flow-through conditions, or where appropriate, semi-static conditions. The test is begun by placing fertilized eggs in the test chambers and is continued at least until all the control fish are free-feeding. Lethal and sub-lethal effects are assessed and compared with control values to determine the lowest observed effect concentration and hence the no observed effect concentration [1].

Developmental stages in the life cycles of fish are relatively sensitive to toxicants. Early life stage (ELS) tests, in which fish are exposed during embryogenesis and larval development, are an essential element in hazard assessment as they have a high predictive power for life-cycle toxicity [2].

3.4.Endpoint units:

µmol/L

3.5.Dependent variable:

log NOEC

3.6.Experimental protocol:

Reconstituted water was used in all tests, prepared from groundwater obtained from a locality near Linschoten, to which several salts were added. This type of reconstituted water has been found to be suitable for breeding a variety of aquatic species. Hardness, expressed as CaCO3, was about 210 mg/l. The mean dissolved oxygen (D.O.) concentration was 7.7 mg/l. The lowest D.O. concentration measured during the tests was 5.1 mg/l. The equilibrium pH of the medium, after aeration, varied from 8.0-8.2. The lowest and highest pH values measured during the tests were 7.4 and 8.4, respectively. The concentrations of the macronutrients were as follows: Na+ (1.19 mmol/l), K+ (0.20 mmol/l), Ca2+ (1.36 mmol/l), Mg2+ (0.73 mmol/l), Cl- (2.72 mmol/l), SO42- (0.73 mmol/l) and HCO3- (1.39 mmol/l). The

groundwater contained several trace elements at concentrations < < 1 mg/l.

Fertilized eggs of zebra fish (Brachydanio rerio) in the blastula stage were obtained from a stock culture at the TNO laboratory of 50-100 eggs (<6 h after spawning) into l-liter glass test vessels filled with l-liter test solution. After 1 day all non-viable eggs were removed and the number of viable eggs was reduced to a maximum of 40 per concentration. In case the number of viable eggs in the controls fell below 25 after 48 h, the test was discarded. The embryolarval stages were exposed in a semistatic manner to 7-8 toxicant concentrations and a control for a period of 28 days. Upon completion of hatching (4-5 days), the fry were transferred into two vessels per concentration. The fry were fed equal amounts of the rotifer Brachionus rubens, obtained from a laboratory culture. After 7 days this food was supplemented by 48-h old nauplii of the brine shrimp Artemia salina. The nauplii were enriched with Selco, a commercial concentrate for nutritional enrichment of live food for fish.

The toxicity tests were carried out in a constant-temperature room at 24 ± 2°C and a photoperiod of 12 h. Dead eggs and larvae were counted and removed daily. At the end of the test period the surviving fish were anesthetized in buffered tricaine methane sulphonate (MS 222, Sandoz, Basel) for final length measurements. The number of microscopically malformed fish was determined under a microscope (magnification 30x).

The ratio between the concentrations was 1.8. The test solutions were renewed 3 times a week. In the tests with the aniline derivatives, the test solutions were gently aerated, in the chlorobenzene tests they were not. In several instances dimethyl sulphoxide (DMSO) was used as solvent for the test compounds. DMSO concentrations were kept below 100 µl/l. The effects of DMSO were verified in solvent control experiments, pH and 02 concentrations were measured at regular intervals.

The actual concentrations of the test compounds were verified before and after renewal of the test solutions during the experiments. Aniline and the chlorinated anilines were analyzed by direct injection of the water samples into a Waters 710B HPLC equipped with a Waters 6000A pump and a Kratos Spectroflow UV detector at 220/240 nm. A Guard 30-40 µm precolumn (Chrompack) and a Vydac 201 TPB 5 µm 100 mm × 3 mm column (Chrompack) were used. HPLC-water + D4 reagent (Waters) and methanol HPLC-grade (Rathburn) was used as eluent. The concentrations of monoCB, 4-chlorotoluene and 1,4-diCB were analyzed in the same manner but detection took place at 220 nm, whereas a mixture of 50% HPLC-water and 50% methanol was used as eluent. 1,2,3-triCB, 1,2,3,4-tetraCB and pentaCB were analyzed on a gas-chromatograph fitted with a 15 m x 0.25 mm DB-1 column and an electron-capture detector. These analyses were carried out on toluene (1:1) extracts from the water samples [3].

The LC50 and 95% confidence limits (C.L.) were calculated according to Kooyman [4]. If a test yielded concentrations without partial kills, the geometric mean of the 0 and 100% effect concentrations was taken as the LC50 and binomial confidence limits were calculated [5]. In order to calculate the no observed lethal concentration (NOLC: the highest concentration tested without significant effects on survival) and no observed effect concentration (NOEC: the highest concentration tested without significant effects on survival, hatching and growth), a two-stage approach was applied to exclude any possible effects of size-selective mortality. First the NOLC was determined. Differences in mean survival in the experimental concentrations were tested against the blank control by means of a $\chi 2$ test [6]. Differences in mean length between treatments and blank control were tested using procedures described by Williams [7, 8], after verifying the differences between blank and solvent controls. The Williams' test was applied only to those concentrations which were equal to or below the NOLC. Differences were considered to be significant at $\alpha = 0.05$.

3.7.Endpoint data quality and variability:

Experimental data from a single lab and series of experiments was used.

Statistics:
max value: 3.517
min value: -1.434
standard deviation: 1.189
skewness: 0.069

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

2D and 3D regression-based QSAR

4.2.Explicit algorithm:

multilinear regression QSAR
multilinear regression QSAR derived with BMLR (Best Multiple Linear Regression) method

log NOEC = 4.034 -77.799×FHACA Fractional HACA (HACA/TMSA) (AM1)
+1.937×Polarity parameter (AM1) / square distance
-1.114×The octanol/water partition coefficient (calc.)

4.3.Descriptors in the model:

[1]FHACA Fractional HACA (HACA/TMSA) (AM1) - relation of hydrogen bond acceptors' surface area to the total molecular surface area as obtained from semi-empirical calculation
[2]Polarity parameter (AM1) / square distance [au/ Å2] difference of maximum positive and negative partial charges (from AM1 calculations) divided by their distance square

[3]The octanol/water partition coefficient (calc.) - logKow based on atom contribution

4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules (one-parameter equations: Fisher criterion and R2 over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold),

(two-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 AM1 calculation. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

QSARModel 5.0.0

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:

8.66(6), (26 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:

a) by chemical identity: structurally heterogeneous organic compounds, aliphatic, cyclic and aromatic hydrocarbons and carbonyl compounds, ethers, halogenoderivatives

b) by descriptor value range: The model is suitable for compounds that have the descriptors

in the following range:

FHACA Fractional HACA (HACA/TMSA) (AM1) 0 .. 0.0174

Polarity parameter (AM1) / square distance -0.12 ..0.506

The octanol/water partition coefficient (calc.) -0.832..6.43

5.2.Method used to assess the applicability domain:

Range of descriptor values in training set with ±30% confidence. Descriptor values must fall between maximal and minimal descriptor values of training set ±30%.

5.3.Software name and version for applicability domain assessment:

QSARModel 5.0.0

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:

See 5.1

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:Yes
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:

26 data points
6 negative values
20 positive values

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

R2 = 0.953 (Correlation coefficient)
s2 = 0.293 (Standard error of the estimate)
F = 150.1 (Fisher function)

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

R2cv = 0.941 (Cross-validated correlation coefficient)

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

n/a

6.10.Robustness - Statistics obtained by Y-scrambling:

n/a

6.11.Robustness - Statistics obtained by bootstrap:

n/a

6.12.Robustness - Statistics obtained by other methods:

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:Yes
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:

2 data points
2 positive values
0 negative values

7.6.Experimental design of test set:

From sorted data each 10th was subjected to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R2 = 0.999 (Correlation coefficient)

7.8.Predictivity - Assessment of the external validation set:

Descriptor value range (all in range of applicability domain):
FHACA Fractional HACA (HACA/TMSA) (AM1) 0 .. 0
Polarity parameter (AM1) / square distance 0.056 ..0.006
The octanol/water partition coefficient (calc.) 2.186..3.43

7.9.Comments on the external validation of the model:

The validation correlation coefficient (R2) for the test set is very high. However, limited size of testing set (and dataset in general) hinderes predictivity assessment

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The main part of the variance of the endpoint values is covered by the octanol/water partition coefficient. The negative coefficient indicates that compounds of higher logKow values (more hydrophobic) show increased toxicity. The hydrogen bond acceptor and dipolarity properties of molecules are described by the other two descriptors (FHACA Fractional HACA (HACA/TMSA) (AM1), Polarity parameter (AM1) / square distance) and appear as minor correction terms.

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:

Interpretation in general agreement with literature [3].

9.Miscellaneous information

9.1.Comments:

The modeling of toxicological properties is an extremely important problem. No empirical toxicological data are available for most chemicals, and the growing new ones must be evaluated or, at least

estimated. Thus, reliable methods to predict environmental toxicity are required.

9.2.Bibliography:

[1]Fish, early-life stage toxicity test OECD TG 210, 1992.
[2]McKim, J. M., Early life stage toxicity tests. In: Fundamentals of aquatic toxicology, edited by G. M. Rand and S. R. Petrocelli, Hemisphere, Washington, 1985, pp. 58-95.
[3]Van Leeuwen C. J., Adema D. M. M. and Hermens J. Quantitative structure-activity relationships for fish early life stage toxicity. Aquatic Toxicology, 1990 (16), 321-334.
[4]Kooyman, S. A. L. M. Parametric analyses of mortality rates in bioassays. Water Res., 1981 (15), 107-119.
[5]Stephan, C. E. Methods for calculating an LC50. In: Aquatic toxicology and hazard evaluation, edited by F. L. Mayer and J. L. Hamelink, A.S.T.M., 1977, STP 634, pp. 65-84.
[6]Sokal, R. R. and Rohlf F. J. Biometry, the principles and practice of statistics in biological research. W. H. Freeman and Co., San Francisco, 1981, 859 pp.
[7]Williams, D. A. A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics, 1971 (27), 103-117.
[8]Williams, D. A. The comparison of several dose levels with a zero dose control. Biometrics, 1972 (28), 519-531.

9.3.Supporting information:

Training set(s)Test set(s)Supporting information

Karelson Arkivoc 2008	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2008.pdf
Karelson Arkivoc 2009	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2009.pdf

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

10.4.Comments: