	QMRF identifier (JRC Inventory):**Q13-16-0059**
	QMRF Title:**TerraQSAR – LOGP**
	Printing Date:**Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

TerraQSAR – LOGP

1.2.Other related models:

TerraQSAR - Daphnia; 48-hr LC50 for Daphnia magna immobilization

TerraQSAR - E2-RBA; estrogen receptor binding affinity (RBA), relative to 17 beta-estradiol (E2)

TerraQSAR - FHM; 96-hr LC50 for fathead minnow (Pimephales promelas)

TerraQSAR - OMAR; mouse & rat oral LD50

TerraQSAR - RMIV; rat & mouse intravenous LD50

TerraQSAR - SKIN; skin irritation potential

1.3.Software coding the model:

TerraBase

Dr. K. Kaiser

http://www.terrabase-inc.com

2.General information

2.1.Date of QMRF:

January 2010

2.2.QMRF author(s) and contact details:

Dr. K. Kaiser TerraBase Inc. mail@terrabase-inc.com http://www.terrabase-inc. com

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Dr. K. Kaiser TerraBase Inc. mail@terrabase-inc.com http://www.terrabaseinc. com

2.6.Date of model development and/or publication:

2004 - 2008

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

[1]Kaiser KLE (2003). Neural networks for effect prediction in environmental and health issues using large datasets. QSAR & Combinatorial Science 22, 185-190 .

[2]Kaiser KLE & Niculescu SP (2001). Modeling acute toxicity of chemicals to Daphnia magna: a probabilistic neural network aaproach. Environmental Toxicology and Chemistry 20, 420-431.

[3]Kaiser KLE, Niculescu SP & Schultz TW (2002). Probabilistic neural network modeling of the toxicity of chemicals to Tetrahymena pyriformis with molecular fragment descriptors. SAR and QSAR in Environmental Research 13, 57-67.

2.8.Availability of information about the model:

Model is proprietary. Algorithm not provided. Training set provided.

Further information in the software user manual, available at

http://www.terrabase-inc.com////TerraQSAR-LOGP.pdf

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

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

None -physicochemical effect

3.2.Endpoint:

1.Physical Chemical Properties 1.6.Octanol-water partition coefficient (Kow)

3.3.Comment on endpoint:

Octanol/water partition coefficient, commonly known as log(Kow), or logP

3.4.Endpoint units:

logarithmic untis

3.5.Dependent variable:

log(Kow), or logP

3.6.Experimental protocol:

3.7.Endpoint data quality and variability:

Various literature data, all from recognized sources, such as Pubchem, or directly from primary publications.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Neural Network

4.2.Explicit algorithm:

Neural Network

Probabilistic Neural Network

The "probabilistic neural network" is substantially different from other neural networks (such as backpropagation, etc.) by having a unique, strictly input-dependent outcome. The same input will result in identical outcomes, on all computation systems, every time.

The algorithm has been described in the literature, see, for example Specht (1988): Probabilistic networks for classification, mapping, or associative memory, ICCN Conference. See also bibliography section of the Manual.

4.3.Descriptors in the model:

4.4.Descriptor selection:

By expert knowledge.

Structural components, including Acidity, Aliphatic ring, Aromatic ring, Atom, Bond, Group, Hydrophobicity, Ionization, Polarity, Reactivity, Stereo, and Weight fragments

4.5.Algorithm and descriptor generation:

4.6.Software name and version for descriptor generation:

4.7.Chemicals/Descriptors ratio:

approx 20 (4010 chemicals/ >200 descriptor)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The probabilistic neural network does not require the external definition of an applicability domain; it is inherent in the structural

variety of the training set of >4000 compounds. Therefore, the probabilistic neural network does not have a limitation of its applicability domain, other than given by the training set condition of "carbon-containing" chemical structures. In other words, the TerraQSAR - LOGP program can be used for any substance fulfilling that condition.

5.2.Method used to assess the applicability domain:

We view the term "applicability domain" as a somewhat vague and certainly not very practical concept (see, for example, EHP 112 (2): correspondence (2004). As we give the entire training set, the user can determine for him/herself as to whether his/her requirements as to applicability domain are being satisfied or not.

5.3.Software name and version for applicability domain assessment:

5.4.Limits of applicability:

Carbon-containing structures; see also the chapter 5.1 and 5.2 for additional comments.

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: No

Chemical Name: No

Smiles: Yes

Formula: No

INChI: No

MOL file: No

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:

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

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

LOO-RMSE = 0.11

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

6.10.Robustness - Statistics obtained by Y-scrambling:

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

No

7.2.Available information for the external validation set:

CAS RN: No

Chemical Name: No

Smiles: No

Formula: No

INChI: No

MOL file: No

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

No

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

No

7.5.Other information about the external validation set:

7.6.Experimental design of test set:

7.7.Predictivity - Statistics obtained by external validation:

7.8.Predictivity - Assessment of the external validation set:

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:

Structural components as described above and in the Manual

8.2.A priori or a posteriori mechanistic interpretation:

Many of the fragments used are those commonly used in other models as well; a few are proprietary. For most of the descriptors, mechanistic interpretations have been described in various publications. While there may well be valid mechanistic interpretations for some or all of our model parameters, we are not prepared to speculate on such.

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

TerraQSAR - LOGP Manual http://www.terrabase-inc.com////TerraQSAR-LOGP.pdf

9.3.Supporting information:

Terra QSARTraining set_4010.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-16-0059/attachment/A747

Test set(s)Supporting information

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q13-16-0059

10.2.Publication date:

2013-07-02

10.3.Keywords:

TerraQSAR;logP;octanol-water partition coefficient;neural network;

10.4.Comments:

former Q18-32-33-245