| | |
|---|---|
| | *QMRF identifier (JRC Inventory):* **Q13-44-0058** |
| | *QMRF Title:* **Nonlinear QSAR: artificial neural network for dermal irritation** |
| | *Printing Date:* **Dec 11, 2019** |
| | |

## 1.QSAR identifier

### 1.1.QSAR identifier (title):

Nonlinear QSAR: artificial neural network for dermal irritation

### 1.2.Other related models:

http://reachqsar.com/

### 1.3.Software coding the model:

QSARModel 3.3.8

Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

Statistica 7

StatSoft Ltd.

http://www.statsoft.com/

## 2.General information

### 2.1.Date of QMRF:

10.10.2010

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

Molcode model development team 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:

12.04.2010

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

[1]Katritzky A R, Dobchev DA, Fara DC, Hur E, Tämm K, Kuruncz L, Karelson M, Varnek A & Solov'ev VP (2006). Skin Permeation Rate as a Function of Chemical Structure. Journal of Medicinal Chemistry 49, 3305-3314.

[2]Karelson M, Dobchev DA, Kulshyn OV & Katritzky A (2006). Neural Networks Convergence Using Physicochemical Data. Journal of Chemical Information and Modeling 46, 1891- 1897.

[3]Statistica 7 www.statsoft.com

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

Selection, 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:

Rabbit

### 3.2.Endpoint:

4.Human Health Effects 4.4.Skin irritation /corrosion

### 3.3.Comment on endpoint:

Dermal irritation is the production of reversible inflammatory changes in irritation skin following the application of a substance. The skin irritation potential is described by the Primary Irritation Index (PII), calculated from erythema and oedema grades based on experimental rabbits. The maximum PII is 8 and the minimum is 0. The grading scale for irritant effects on rabbit skin were originally proposed by Draize and adopted by the OECD (Test Guideline 404) and the US and EU regulatory agencies [ref 1, sect 9.2].

The PII can be calculated as:

PII = [SUM(Erythema 24/48/72 h)+SUM(Oedema24/48/72 h)] (3 x no. animals)

where Erythema is redness of skin produced by vascular congestion or increased perfusion And Oedema is the presence of abnormally large amounts of fluid in the intercellular tissue spaces of the epidermis, dermis or subcutaneous tissues.

### 3.4.Endpoint units:

### 3.5.Dependent variable:

Primary Irritation Index (PII)

### 3.6.Experimental protocol:

All 286 data used in this report were obtained from in vivo rabbit skin irritation test that were used to assess the potential of materials to cause skin irritancy or corrosion in man, and to meet regulatory requirements which require classification and appropriate labeling of a material if it is believed to be potential irritant or corrosive. All chemicals were tested applying a volume or weight of 0.5ml or 0.5g undiluted, except where an alternative weight or concentration was needed. Exposure time for each test was 4 hours [ref 1, sect 9.2].

### 3.7.Endpoint data quality and variability:

The 286 chemicals selected were readily available at high and consistent purity and are expected to be stable on storage. They have been tested undiluted in in vivo studies, excepting those chemical where high concentrations of the substance could be expected to cause sever effects. The invivo data were generated in 1981 in studies carried out according to OECD Test Guideline 404 and following the principles of Good Laboratory Practice. The data presented were obtained from tests normally using at least three rabbits involving application of 0.5 ml (or 0.5g) to the flank under semi-occlusive patches and in which observations were made at least 24, 48 and 72 hours [ref 1, sect 9.2].

## 4.Defining the algorithm - OECD Principle 2

**4.1.Type of model:**

Neural network

**4.2.Explicit algorithm:**

Neural network

Nonlinear QSAR: backpropagation Neural Network (Multilayer Perceptron) regression

The algorithm is based on regression neural network predictor with structure 9-8-6-1.

**4.3.Descriptors in the model:**

[1]count of H-acceptor sites (AM1)

[2]HBCA H-bonding charged surface area (AM1)

[3]FHACA Fractional HACA (HACA/TMSA) (AM1)

[4]Average atom weight

[5]min(#HA, #HD) (AM1)

[6]HACA-2 (AM1)

[7]Number of O atoms

[8]Difference (Pos - Neg) in Charged Surface Areas (Zefirov)

[9]Negatively Charged Part of Charged Surface Area (AM1)

**4.4.Descriptor selection:**

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules as F statistic and p. The first highest F (low p) descriptors (9) were selected from the whole set of descriptors. These 9 descriptors were used as inputs to the network. 29 networks with different structures were tested in order to find the best ANN with lowest RMS (root-mean-squared error) and highest correct predictions (for training, selection and test sets). Then 245 epochs were used to train the final network with architecture depicted in 4.2. Optimization of the weights was performed with Levenberg-Marquardt algorithm encoded in the backpropagation scheme using linear and hyperbolic activation functions.

**4.5.Algorithm and descriptor generation:**

All descriptors were generated using QSARModel on structure optimized by AM1 semiempirical quantum mechanical model. The final structure were optimized by mopac6 implemented in QSARModel. Keywords used for opyimizations were: AM1 EF GNORM=0.05 BONDS PI POLAR ENPART NOINTER PRECISE. The final descriptors were selected as denoted in 4.4 as well as descriptors with small variances less than 10 $e^{-5}$were discarded from the total pool.

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

QSARModel 3.3.8

http://www.molcode.com

**4.7.Chemicals/Descriptors ratio:**

16.2 ( 146 chemicals / 9 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 descriptor value range (between min and max values): The model is suitable for compounds that have the descriptors in the following range augmented with the confidence in 5.2:

Desc ID

See 4.3: 1 2 3 4 5 6 7 8 9

Min: 0.000 0.000 0.000 4.588 0.000 0.000 0.000 -243.84 1.705

Max: 4.000 92.304 0.209 27.639 4.000 6.156 6.000 952.967 140.451

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

Presence of functional groups in structures (ethers, esters, amides, halides, aromatic, aliphatic functional groups etc)

Range of descriptor values in training set with ±30% confidence

Descriptor values must fall between maximal and minimal descriptor values (see5.1) of training set ±30%.

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

QSARModel 3.3.8

http://www.molcode.com

**5.4.Limits of applicability:**

See 5.1, 5.2

## 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:**

Data points: 146

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

Standardization and normalization of the inputs by taking into account the mean and standard deviation

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

Training PII; Selection PII; Test PII

Data Mean: 2.348; 3.129; 2.417

Data SD: 2.040; 2.512; 1.610

Error Mean: -0.019; -0.009; -0.222

Error SD: 1.185; 2.781; 1.390

Abs E. Mean: 0.845; 1.903; 1.112

S.D. Ratio: 0.581; 1.107; 0.864

Correlation: 0.814; 0.628; 0.590

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

See 6.7

**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:**

RMS(Training)=0.14814; RMS (Selection)=0.347624; RMS(Test)=0.176003

In this ANN 2 sets of randomly chosen (20) data to test the network – selection set and test set, See also 6.7

## 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:**

The method used two validation sets: selection (20) and test (20)

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

Randomly selected 20 selection and 20 test data points

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

See 6.7 and 6.12

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

The descriptors for the test set are in the limit of applicability, see 6.7 and 6.12. We have limited ourselves to select two auxiliary sets to train the network and to test it externally on the test set. Thus more than 1.5 of the datapoints were selected for these two sets divided by 2. One of the main purposes of the ANN model also to be applicable for diverse compounds for future predictions, thus we wtried to keep the training set as large as possible and to select the validation and test sets with significant data points.

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

Overall predictions for the selection set (used to stop the ANN training and not to overfit it) and the test set (used to test the external prediction of the net after training) are significant according to the RMS error and the standard deviation ratio (S.D.Ratio); see 6.7 and 6.12.

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

### 8.1.Mechanistic basis of the model:

The complex nature of the ANN model does not allow direct interpretation of the descriptors in relation to the modelled property. However, it can be noted that descriptors related to the hydrogen bonding ability and the charged surface areas of the molecules are mainly present. The reactivity of the compounds with the epidermis depends also on charged surface areas of the compounds (which are the most reactive sides). Several authors have confirmed the reactivity related with the charged surfaces and also the LUMO and HOMO descriptors [ref 2,3; sect 9.2]. It can be roughly estimated that the PII increases with increasing (slight negative correlation between the descriptors) count of H-acceptor sites (AM1), HBCA H-bonding charged surface area (AM1), and FHACA Fractional HACA (HACA/TMSA) (AM1).

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

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

## 9.Miscellaneous information

### 9.1.Comments:

Supporting information for: training set(s), selection set(s), test set(s).

The 9-8-6-1.snn file includes the ANN model; the user must have Statistica 7 or higher with ANN modules to make predictions.

### 9.2.Bibliography:

[1]ECETOC Technical Report No 66. Skin Irritation and Corrosion: Reference Chemicals Data Bank. March 1995

[2]Kodithala K, Hopfinger AJ, Thompson ED & Robinson MK (2002). Prediction of skin irritation from organic chemicals using membrane-interaction QSAR analysis. Toxicological Sciences 66, 336–346.

[3]Hayashi M, Nakamura Y, Higashi K, Kato H, Kishida F & Kaneko H (1999). A quantitative structure-Activity relationship study of the skin irritation potential of phenols. Toxicology in Vitro 13, 915-922.

### 9.3.Supporting information:

| | |
|---|---|
| Dermal_Irritation_PII_training_ 146.sdf | http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-44-0058/attachment/A743 |
| Dermal_Irritation_PII_test_20.sdf | http://qsardb.jrc.ec.europa.eu/qmrffile:///C:\Documents and Settings\casadja\My Documents\a_QSAR\a_QMRFdb\model submissions\7 review completed\Q17-22-1-332_Nonlinear ANN QSAR Model for Dermal Irritation\Dermal_Irritation_PII_test_20.sdf |
| Dermal_Irritation_PII_selection_20.sdf | http://qsardb.jrc.ec.europa.eu/qmrffile:///C:\Documents and Settings\casadja\My Documents\a_QSAR\a_QMRFdb\model submissions\7 review completed\Q17-22-1-332_Nonlinear ANN QSAR Model for Dermal Irritation\Dermal_Irritation_PII_selection_20.sdf |

**Supporting information**

## 10.Summary (JRC QSAR Model Database)

### 10.1.QMRF number:

Q13-44-0058

### 10.2.Publication date:

2013-07-01

### 10.3.Keywords:

skin irritation;Primary Irritation Index;PII;Draize;Molcode;

### 10.4.Comments:

former Q17-22-1-332