	QMRF identifier (JRC Inventory): **Q17-44-0003**
	QMRF Title: **BIOVIA toxicity prediction model – skin irritancy (moderate vs severe)**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

BIOVIA toxicity prediction model – skin irritancy (moderate vs severe)

1.2.Other related models:

BIOVIA toxicity prediction model – skin irritancy (none vs irritant)

BIOVIA toxicity prediction model – skin irritancy (mild vs moderate/severe)

1.3.Software coding the model:

BIOVIA Discovery Studio v4.5

Dassault Systèmes, BIOVIA Corp., 5005 Wateridge Vista Drive, San Diego, CA92121, USA

http://www.3dsbiovia.com

2.General information

2.1.Date of QMRF:

16 January 2017

2.2.QMRF author(s) and contact details:

Deqiang Zhang Dassault Systemes, BIOVIA Corp. 5005 Wateridge Vista Drive, San Diego, CA 92121, USA Deqiang.Zhang@3ds.com http://www.3dsbiovia.com

2.3.Date of QMRF update(s):

N/A

2.4.QMRF update(s):

N/A

2.5.Model developer(s) and contact details:

Deqiang Zhang Dassault Systemes, BIOVIA Corp. 5005 Wateridge Vista Drive, San Diego, CA 92121, USA Deqiang.Zhang@3ds.com http://www.3dsbiovia.com

2.6.Date of model development and/or publication:

2015

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

BIOVIA Discovery Studio v4.5 http://www.3dsbiovia.com/products/discovery-studio/

2.8.Availability of information about the model:

The model is proprietary (available as a commercial product), but the algorithm is public. The training set is also proprietary, however, it is embedded with the model and can be retrieved with similarity search when a prediction is conducted. No external test is conducted except cross-validation.

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

None

3.Defining the endpoint - OECD Principle 1

3.1.Species:

N/A

3.2.Endpoint:

4.Human Health Effects 4.4.Skin irritation /corrosion

3.3.Comment on endpoint:

This model was trained using 601 samples showing moderate or severe irritancy out of 1277 samples with Rabbit Skin Irritancy test result collected from literature data. The sources of data include:

1. Research Institute of Fragrance Materials
2. U.S. Army Environmental Hygiene Agency
3. Various other open literature sources

The classification of skin irritants is as follows:

Category Draize Scale (Draize et al., 1944) * Primary Irritation Index (Smyth et al., 1944) *

None 0 0

Mild 1 >0 - 2.0

Moderate 2 2.1 - 5.0

Severe 3 > 5.0

* Draize, Woodard, and Calvery, J. Pharmaco. Exp. Ther. 82, 1944, pp. 377-390.

* Smyth and Carpenter, J. Ind. Hyg. Toxicol. 26, 1944, pp. 269-273.

* Primary (Dermal) Irritation Index is also known as 10-point score.

3.4.Endpoint units:

Dimensionless - Yes/No Binary Classification

3.5.Dependent variable:

Classification as moderate irritant or severe irritant

3.6.Experimental protocol:

The test protocol is the in vivo acute skin irritation method using the albino rabbit, outlined in OECD Guidelines for the Testing of Chemicals, Section 4, Test No. 404: Acute Dermal Irritation/Corrosion, available online at

3.7.Endpoint data quality and variability:

All the data were collected from literature data. The sources of data include:

Research Institute of Fragrance Materials

U.S. Army Environmental Hygiene Agency

Various other open literature sources

The classification of skin irritants is as follows:

Category Draize Scale (Draize et al., 1944) * Primary Irritation Index (Smyth et al., 1944) *

None 0 0

Mild 1 >0 - 2.0

Moderate 2 2.1 - 5.0

Severe 3 > 5.0

* Draize, Woodard, and Calvery, J. Pharmaco. Exp. Ther. 82, 1944, pp. 377-390.

* Smyth and Carpenter, J. Ind. Hyg. Toxicol. 26, 1944, pp. 269-273.

* Primary (Dermal) Irritation Index is also known as 10-point score.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR model derived from Bayesian binary classification

4.2.Explicit algorithm:

Bayesian Classification

A modified Bayesian learning method is used. The algorithm is described in Xia X, Maliski EG, Gallant P & Rogers D(2004). Journal of Medicinal Chemistry. 47(18) 4463- 4470

Pcorr(Active|F) = (A + P(Active)*K)/(B + K).

(For K = 1/P(Active), this is the Laplacian correction.)

4.3.Descriptors in the model:

[1]ALogP unitless The calculated partition-coefficient of a compound between 1-octanol and water

[2]Molecular_Weight gram/mole The calculated molecular weight by summing the average atomic weight of all the atoms in the molecule.

[3]Num_H_Donors unitless Number of hydrogen bond donors.

[4]Num_H_Acceptors unitless Number of hydrogen bond acceptors in the molecule.

[5]Num_RotatableBonds unitless Number of rotatable bonds in the molecule.

[6]Molecular_PolarSurfaceArea Angstrom^3 The polar surface area of the molecule.

[7]SCFP_12 unitless Extended-connectivity SYBYL atom type fingerprint with a maximum length of 12 bonds

[8]Molecular_FractionalPolarSurfaceArea Unitless The fraction of the polar surface area over the total molecular surface area.

4.4.Descriptor selection:

A pool of most commonly used descriptors (ALogP, Molecule_Weight, Num_H_Donors, Num_H_Acceptors, Molecular_FractionPolarSurfaceArea, ECFP_2, ECFP_4, ECFP_6, ECFP_8, ECFP_10, ECFP_12, FCFP_2, FCFP_4, FCFP_6, FCFP_8, FCFP_10, FCFP_12, SCFP_2, SCFP_4, SCFP_6, SCFP_8, SCFP_10, SCFP_12) were selected randomly to build models. The model with the best leave-one-out cross-validated ROC score is selected to build the final model. In addition, the Bayesian model has a built-in mechanism to select the most statistically-significant descriptors.

4.5.Algorithm and descriptor generation:

(1) The ALogP is the Ghose/Crippen group-contribution estimate for LogP, where P is the relative solubility of a compound in octanol versus water. See Ghose, A.K., Viswanadhan, V.N., and Wendoloski, J.J., "Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragment Methods: An Analysis of AlogP and CLogP Methods." J. Phys. Chem. A, 1998, 102, 3762-3772.

(2) Molecular weight is calculated using the atomic weights of the individual atoms in the molecule.

(3) Hydrogen bond acceptors are defined as heteroatoms (O, N, S, or P) with one or more lone pairs, excluding atoms with positive formal charges, amide and pyrrole-type nitrogens, and aromatic oxygen and

sulfur atoms in heterocyclic rings.
(4) Hydrogen bond donors are defined as heteroatoms (O, N, S, or P) with one or more attached hydrogen atoms.
(5) Molecular_PolarSurfaceArea and Molecular_FractionPolarSurfaceArea are calculated from the polar surface area and total surface area using a 2D approximation to each molecule.
(6) The fingerprint generation method is based on one of the original algorithms in computational organic chemistry called the Morgan algorithm. The goal of the Morgan algorithm is to assign a unique identity to each atom in a molecule so that a molecule can be described in a way that is invariant to the original numbering of atoms. The algorithm has two parts: the assignment of an initial code to each atom, and an iterative part in which each atom code is updated to reflect the codes of each atom's neighbors.
SCFP_12 is calculated by first assigning atom types (SCFP_0) using SYBYL atom types, and an n iterative process is used to generate features that represent each atom in progressively larger structural neighborhoods. After each iteration, the new feature codes for the atoms are added to the set of features from all previous steps. The process completes when the desired size is reached and the set of all features is returned as the fingerprint.

4.6.Software name and version for descriptor generation:

Dassult Systemes BIOVIA Pipeline Pilot Server
U.S. 6am -4pm (Pacific Time) Toll Free: 1-800-756- 4674 Tel: (858) 799-5509 support@accelrys.com U.K. 9:00 to 16:00 (UK time) Tel: +44 1223 228822 UK local rate: +44 845 741 3375 Central Europe 9:00 to 16:00 (Central European time) Switzerland: Tel: +41 61 588 0480 Germany: Tel: +49 221 8282 9020 support@accelrys.com Japan 10:00 to 17:00 (Tokyo time) Toll Free: 0120-712655 Tel: +81 3 4321 3906 support-japan@accelrys.com http://accelrys.com/products/collaborative-science/biovia-pipeline-pilot/

4.7.Chemicals/Descriptors ratio:

Number of chemicals = 601
Number of descriptors = 8
Chemicals/Descriptors = 75

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain of the model is defined by the range of descriptors of training set chemicals. The applicability domain is only a qualitative measure on how reliable the prediction is. There is no quantitative measure on how reliable the prediction is.

5.2.Method used to assess the applicability domain:

If a continuous descriptor is out of range of the training set, a warning is issued for the input compound. For the fingerprint descriptors, if a new feature not seen in the training set is found, a warning message is issued for that feature.

5.3.Software name and version for applicability domain assessment:

Dassult Systemes BIOVIA Pipeline Pilot Server

U.S. 6am -4pm (Pacific Time) Toll Free: 1-800-756- 4674 Tel: (858) 799-5509 support@accelrys.com U.K. 9:00 to 16:00 (UK time) Tel: +44 1223 228822 UK local rate: +44 845 741 3375 support@accelrys.com Central Europe 9:00 to 16:00 (Central European time) Switzerland: Tel: +41 61 588 0480 Germany: Tel: +49 221 8282 9020 support@accelrys.com Japan 10:00 to 17:00 (Tokyo time) Toll Free: 0120-712655 Tel: +81 3 4321 3906 support-japan@accelrys.com http://accelrys.com/products/collaborative-science/biovia-pipeline-pilot/

5.4.Limits of applicability:

Property Min Max Mean Std. Dev.

ALogP -6.143 14.177 2.592 2.0064

Molecular_Weight 30.026 697.69 185.16 81.969

Num_H_Donors 0 8 0.42928 0.75792

Num_H_Acceptors 0 13 1.9218 1.3805

Num_RotatableBonds 0 27 4.3245 4.1988

Molecular_PolarSurfaceArea 0 252.19 30.722 25.11

Molecular_FractionalPolarSurfaceArea 0 0.909 0.15504 0.1172

SCFP_12 N/A N/A N/A N/A

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: Yes

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:

The data used to train the model consisted of 601 samples showing either moderate or severe irritancy in the test. 266 of them are in the positive category (severe). The training set is proprietary, however, it is embedded with the model and can be retrieved with similarity search when a prediction is conducted.

6.6.Pre-processing of data before modelling:

None

6.7.Statistics for goodness-of-fit:

N/A

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

ROC score=0.819 (LOO)

True Positive = 207

False Negative = 59

False Positive = 78

True Negative = 257

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:

N/A

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:

No data were reserved for external validation purpose because the sample size is small.

7.6.Experimental design of test set:

N/A

7.7.Predictivity - Statistics obtained by external validation:

N/A

7.8.Predictivity - Assessment of the external validation set:

N/A

7.9.Comments on the external validation of the model:

N/A

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No mechanistic basis is available for the model because the Bayesian method was used.

8.2.A priori or a posteriori mechanistic interpretation:

N/A

8.3.Other information about the mechanistic interpretation:

N/A

9.Miscellaneous information

9.1.Comments:

The model is extensible, i.e., it can be extended by feeding new
training data to create an improved model.

9.2.Bibliography:

Xia X, Maliski EG, Gallant P & Rogers D(2004). Journal of Medicinal Chemistry. 47(18) 4463- 4470
http://pubs.acs.org/doi/full/10.1021/jm0303195

9.3.Supporting information:

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

10.Summary (JRC Inventory)

10.1.QMRF number:

Q17-44-0003

10.2.Publication date:

2017-09-20

10.3.Keywords:

skin irritation;Draize test;BIOVIA;

10.4.Comments: