	QMRF identifier (JRC Inventory):**Q17-49-0045**
	QMRF Title:**BIOVIA toxicity prediction model –moderate vs severe eye irritant**
	Printing Date:**Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

BIOVIA toxicity prediction model –moderate vs severe eye irritant

1.2.Other related models:

Toxicity Prediction (Extensible) Ocular Irritancy (None vs Irritant)

Toxicity Prediction (Extensible) Ocular Irritancy (Mild vs Moderate/Severe)

1.3.Software coding the model:

BIOVIA Discovery Studio v4.5

Optimize your drug discovery process with a flexible application that delivers predictive science to its required depth.

Dassault Systèmes, BIOVIA Corp., 5005 Wateridge Vista Drive, San Diego, CA92121, USA http://www.3dsbiovia.com

2.General information

2.1.Date of QMRF:

12/5/2015

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, but the algorithm is in public domain. 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:

Rabbit

3.2.Endpoint:

4.Human Health Effects 4.9.Eye irritation/corrosion

3.3.Comment on endpoint:

The ocular irritancy test is traditionally done by using the so called Draize Eye Test. The Draize Test is an acute toxicity test devised in 1944 by Food and Drug Administration (FDA) toxicologists John H. Draize and Jacob M. Spines. Initially used for testing cosmetics, the procedure involves applying 0.5mL or 0.5g of a test substance to the eye or skin of a restrained, conscious animal, and then leaving it for set amount of time before rinsing it out and recording its effects. The animals are observed for up to 14 days for signs of erythema and edema in the skin test, and redness, swelling, discharge, ulceration, hemorrhaging, cloudiness, or blindness in the tested eye. The test subject is commonly an albino rabbit, though other species are used too, including dogs. The animals are euthanized after testing if the test renders irreversible damage to the eye or skin. Animals may be re-used for testing purposes if the product tested causes no permanent damage. Animals are typically reused after a "wash out" period during which all traces of the tested product are allowed to disperse from the test site.

3.4.Endpoint units:

Dimensionless - Yes/No Binary Classification

3.5.Dependent variable:

Classification of an moderate/severe irritant as either moderate or severe. The following classification is used according to Bruner et al. (2004):

None Irritant: Maximum Average Score between 0 and 5;

Mild Irritant: Maximum Average Score between 5 and 15;

Moderate Irritant: Maximum Average Score between 15 and 50;

Severe Irritant: Maximum Average Score between 50 and 110.

The models are comprised of three multiple stages: In the first stage, non-irritants are separate from irritants (including mild, moderate or severe irritants); At the second stage, mild irritants are separated from mild/severe irritants; At the third stage moderate separated from severe irritants.

3.6.Experimental protocol:

The OECD Guidelines for the Testing of Chemicals guideline No. 405 of 2012 contains the guideline gor acute eye irritation/corrosion, available online at

3.7.Endpoint data quality and variability:

The data for this model were collected from 1453 uniform studies selected after critical review of open literature, i.e., all data were from Draize protocol. No further analysis result is available.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

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_FractionalPolarSurfaceArea unitless The fraction of polar surface area over the total molecular surface area.

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

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, Bayesian model has a built-in mechanism to select the most statitstically-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_FractionPolarSurfaceArea is 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_10 is calculated by first assigning a type to each atom using the SYBYL rule (SCFP_0), 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

uilt on the BIOVIA Foundation, Pipeline Pilot enables scientists to rapidly create, test and publish scientific services that automate the process of accessing, analyzing and reporting scientific data, either for the scientist’s personal use or for sharing across the scientific community. Using Pipeline Pilot, scientist, researchers, engineers, and analysts with little or no software development experience can create scientific protocols that can be executed through a variety of interfaces including Accelrys Web Port, other Accelrys solutions such as Accelrys Electronic Lab Notebook, Isentris, Chemical Registration and third-party applications such as Microsoft SharePoint or customer-developed applications. These protocols aggregate and provide immediate access to volumes of disparate research data locked in silos. They automate the scientific analysis of the data and enable researchers to rapidly explore, visualize and report results

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/pipeline-pilot/

4.7.Chemicals/Descriptors ratio:

Number of chemicals = 855

Number of descriptors = 7

Chemicals/Descriptors = 122.1

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 qualititive measure on how reliable the prediction is. There is no quantative 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

Built on the BIOVIA Foundation, Pipeline Pilot enables scientists to rapidly create, test and publish scientific services that automate the process of accessing, analyzing and reporting scientific data, either for the scientist's personal use or for sharing across the scientific community. Using Pipeline Pilot, scientist, researchers, engineers, and analysts with little or no software development experience can create scientific protocols that can be executed through a variety of interfaces including Accelrys Web Port, other Accelrys solutions such as Accelrys Electronic Lab Notebook, Isentris, Chemical Registration and third-party applications such as Microsoft SharePoint or customer-developed applications. These protocols aggregate and provide immediate access to volumes of disparate research data locked in silos. They automate the scientific analysis of the data and enable researchers to rapidly explore, visualize and report results.

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/pipeline-pilot/

5.4.Limits of applicability:

Property Min Max Mean Std. Dev.
ALogP -3.833 16.267 1.7123 1.8962
Molecular_Weight 30.026 925.19 182.4 118.63
Num_H_Donors 0 6 0.78713 0.86468
Num_H_Acceptors 0 17 2.6713 2.2438
Num_RotatableBonds 0 56 4.7918 7.7088
Molecular_FractionalPolarSurfaceArea 0 0.844 0.2227 0.14472
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 855 moderate or severe irritant samples. 530 of them are in the positive category (severe irritant). The training set is proprietary, however, it is emmbedded 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=0.764 (LOO)

True Positive = 377

False Negative = 153

False Positive = 88

True Negative = 237

Sensitivity = 0.811

Specificity = 0.608

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

ROC score = 0.740 (Leave 10% out)

Sensitivity = 0.797

Specificity = 0.892

Concordance = 0.856

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:

Due to the small size of the available data, no data were reserved for external validation purpose.

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:

Features contributing the most from SCFP_12 are included in attachment.

8.2.A priori or a posteriori mechanistic interpretation:

posteriori: these features are selected purely based on their Bayesian score

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:

[1]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

[2]Draize JH, Woodard G, Calvery HO (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. Journal of Pharmacology And Experimental Therapeutics. 82, 377–390 http://jpet.aspetjournals.org/content/82/3/377.citation

[3]Weil CS, Scala RA (1971). Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. Toxicology and Applied Pharmacology. 19(2) 276-360 http://dx.doi.org/10.1016%2F0041-008X%2871%2990112-8

[4]Stokes WS (retrieved 29 June 2009). Preliminary Evaluation of the Underprediction Rate of the In Vivo Dermal Irritation Test Method. US Scientific Advisory Committee on Alternative Toxicological Methods. http://ntp.niehs.nih.gov/index.cfm?objectid=BD09E3A8-F1F6-975E-7C5A023D74150157

9.3.Supporting information:

qmrf512_Features.png	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q17-49-0045/attachment/A1097

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q17-49-0045

10.2.Publication date:

2017-09-27

10.3.Keywords:

rabbit;Draize test;eye irritation;BIOVIA Discovery Studio;

10.4.Comments:

old# Q50-54-55-512