Leadscope Enterprise model for severe skin irritation *in vivo*

## 1. QSAR identifier

### 1.1 QSAR identifier (title)

Leadscope Enterprise model for severe skin irritation *in vivo*, Danish QSAR Group at DTU Food.

### 1.2 Other related models

MultiCASE CASE Ultra model for severe skin irritation *in vivo*, Danish QSAR Group at DTU Food.

SciMatics SciQSAR model for severe skin irritation *in vivo*, Danish QSAR Group at DTU Food.

### 1.3. Software coding the model

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

2. General information

2.1 Date of QMRF

January 2015.

2.2 QMRF author(s) and contact details

QSAR Group at DTU Food;

Danish National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Marianne Dybdahl;

National Food Institute at the Technical University of Denmark;

Sine Abildgaard Rosenberg;

National Food Institute at the Technical University of Denmark;

2.3 Date of QMRF update(s)

2.4 QMRF update(s)

2.5 Model developer(s) and contact details

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Danish QSAR Group at DTU Food;

National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

2.6 Date of model development and/or publication

January 2014.

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

Roberts, G., Myatt, G. J., Johnson, W. P., Cross, K. P., and Blower, P. E. J. (2000) LeadScope: Software for Exploring Large Sets of Screening Data. *Chem. Inf. Comput. Sci.*, 40, 1302-1314.

Cross, K.P., Myatt, G., Yang, C., Fligner, M.A., Verducci, J.S., and Blower, P.E. Jr. (2003) Finding Discriminating Structural Features by Reassembling Common Building Blocks. *J. Med. Chem.*, *46,* 4770-4775.

Valerio, L. G., Yang, C., Arvidson, K. B., and Kruhlak, N. L. (2010) A structural feature-based computational approach for toxicology predictions. *Expert Opin. Drug Metab. Toxicol.*, 6:4, 505-518.

2.8 Availability of information about the model

The training set is non-proprietary and consists of data from the Registry of Toxic Effects of Chemical Substances (RTECS®) database and the Hazardous Substances Data Bank (HSDB) extracted from CHEM-BANK (2000), data from Annex I of Directive 67/548/EEC and expert judgments of certain groups of chemicals. The model algorithm is proprietary from commercial software.

2.9 Availability of another QMRF for exactly the same model

3. Defining the endpoint

3.1 Species

Rabbit (for those chemicals that with experimental data, the remaining chemicals where classified based on experts judgment).

3.2 Endpoint

QMRF 4. Human Health Effects

QMRF 4. 4. Skin irritation /corrosion

3.3 Comment on endpoint

A chemical is characterised as a skin irritant if it produces a non-allergic, inflammatory reaction of the skin (i.e. oedema and erythema) at the site of contact on first exposure. The skin irritating chemical is acting directly on the skin cells and decreases the barrier function of the skin. The damage on the epidermal cells initiates a release of inflammatory mediators that attracts inflammatory cells and an inflammatory reaction in the skin can be seen.
In the training set for this model compounds causing no or mild skin irritation were categorised as mild skin irritants. Compounds causing severe skin irritation were categorised as severe skin irritants. The specific classification into mild and severe skin irritants from the four data sources differs and the criteria can be seen under 3.6.

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Skin irritation *in vivo*, severe or mild.

3.6 Experimental protocol

The training set includes data from RTECS ®, HSDB, Annex I of Directive 67/548/EEC plus expert judgments of certain groups of chemicals. Except for the chemicals classified based on expert judgement chemicals have been tested *in vivo*, according to the experimental protocol for the Draize test (RTECS 2012) or OECD guideline 404 (2002). Detailed descriptions of the experimental protocols can be found in the references.

701 of the training set chemicals were compiled from the RTECS® database (CHEM-BANK 2000). The chemicals in RTECS have been categorised in to three skin irritant categories (RTECS 2012): mild (RTECS code MLD), moderate (RTECS code MOD) and severe (RTECS code SEV). Of the 701 chemicals 291 were categorised as severe skin irritant based on two criteria: 1) they all have the RTECS code SEV, i.e. severe

erythema (beet redness) to slight eschar formation (injuries in depth) and severe oedema (raised more than 1 mm and extending beyond area of exposure), and 2) no requirements on dose or duration of exposure was made. The remaining 410 RTECS chemicals were categorised as mild skin irritants based on the following two criteria: 1) RTECS code MLD, i.e. well defined erythema and slight oedema (edges of area well defined by definite raising), and 2) a requirement of 500 mg and 24 hours exposure.

31 HSDB chemicals (CHEM-BANK 2000) were included in the training set and all were categorised as severe skin irritants, i.e. highly irritating or corrosive, according to HSDB criteria.

56 chemicals from Annex I of Directive 67/548/EEC (EU classifications 1967) were included in the training set as severe skin irritants. These chemicals have EU classifications R34 (causes burns) or R35 (causes severe burns).

The remaining 48 chemicals in the training set were classified in to mild (8) or severe (40) skin irritants based on expert judgments. This group of chemicals consisted of presumably not irritating chemicals that the model was otherwise confused by and where experimental data could not be found (i.e. set to be mild skin irritants), together with some well-known severe skin irritants.

### 3.7 Endpoint data quality and variability

As data originates from multiple sources and different criteria have been used for classification in to mild and severe irritants a certain degree of variability in data is expected.

4. Defining the algorithm

4.1 Type of model

A categorical (Q)SAR model based on structural features and numeric molecular descriptors.

4.2 Explicit algorithm

This is a categorical (Q)SAR model made by use of partial logistic regression (PLR). The model was made using the composite model building option in Leadscope (see 4.4). The composite model consisted of only 1 submodel which uses all the positives and negatives. The specific implementation is proprietary within the Leadscope software.

4.3 Descriptors in the model

structural features,

aLogP,

polar surface area,

number of hydrogen bond donors,

Lipinski score,

number of rotational bonds,

parent atom count,

parent molecular weight,

number of hydrogen bond acceptors

4.4 Descriptor selection

Leadscope Predictive Data Miner is a software program for systematic sub-structural analysis of a chemical using predefined structural features stored in a template library, training set-dependent generated structural features (scaffolds) and calculated molecular descriptors. The feature library contains approximately 27,000 pre-defined structural features and the structural features chosen for the library are motivated by those typically found in small molecules: aromatics, heterocycles, spacer groups, simple substituents. Leadscope allows for the generation of training set-dependent structural features (scaffold generation), and these features can be added to the pre-defined structural features from the library and be included in the descriptor selection process. It is possible in Leadscope to remove redundant structural

features before the descriptor selection process and only use the remaining features in the descriptor selection process. Besides the structural features Leadscope also calculates eight molecular descriptors for each training set structure: the octanol/water partition coefficient (alogP), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), Lipinski score, atom count, parent compound molecular weight, polar surface area (PSA) and rotatable bonds. These eight molecular descriptors are also included in the descriptor selection process.

Leadscope has a default automatic descriptor selection procedure. This procedure selects the top 30% of the descriptors (structural features and molecular descriptors) according to $X^2$-test for a binary variable, or the top and bottom 15% descriptors according to *t*-test for a continuous variable. Leadscope treats numeric property data as ordinal categorical data. If the input data is continuous such as $IC_{50}$ or cLogP data, the user can determine how values are assigned to categories: the number of categories and the cut-off values between categories. (Roberts *et al.*2000).

When developing this model, intermediate models with application of different modelling approaches in Leadscope were tried:

1. 'Single model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors for descriptor selection.
2. 'Single model' using both the Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.
3. 'Single model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation), with subsequent removal of redundant structural features, and calculated molecular descriptors for descriptor selection.
4. 'Composite model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors in the descriptor selection.
5. 'Composite model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.

Based on model performance as measured by a preliminary cross-validation the model developed using approach number 5. was chosen.

For this model scaffolds were generated by Leadscope for the training set structures and added to the Leadscope library of structural features. Descriptors were then automatically selected among the structural features and the eight molecular descriptors.

4.5 Algorithm and descriptor generation

For descriptor generation see 4.4.

After selection of descriptors the Leadscope Predictive Data Miner program performs partial least squares (PLS) regression for a continuous response variable, or partial logistic regression (PLR) for a binary response variable, to build a predictive model. By default the Predictive Data Miner performs leave-one-out or leave-groups-out (in the latter case, the user can specify any number of repetitions and percentage of structures left out) cross-validation on the training set depending on the size of the training set. In the cross-validation

made by Leadscope the descriptors selected for the 'mother model' are used when building the validation submodels and they therefore have a tendency to be overfittet and give overoptimistic validation results.

In this model, because of the categorical outcome in the response variable, PLR was used to build the predictive model. For the single submodel in this composite model 201 descriptors were selected to build the submodel. These include 9 Leadscope calculated molecular descriptors, 70 hierarchy features, 2 dynamic features and 120 scaffolds. The 201 descriptors were distributed on 5 PLS factors.

4.6 Software name and version for descriptor generation

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

4.7 Descriptors/chemicals ratio

In this model the 201 descriptors were used in the single submodel and they were distributed on 5 PLS factors. The training set consists of 836 compounds. The descriptor/chemical ratio is 1:4.2 (201:836).

## 5. Defining Applicability Domain

### 5.1 Description of the applicability domain of the model

The definition of the applicability domain consists of two components; the definition of a structural domain in Leadscope and the in-house further probability refinement algorithm on the output from Leadscope to reach the final applicability domain call.

1. Leadscope

For assessing if a test compound is within the structural applicability domain of a given model Leadscope examines whether the test compound bears enough structural resemblance to the training set compounds used for building the model (i.e. a structural domain analysis). This is done by calculating the distance between the test compound and all compounds in the training set (distance = 1 - similarity). The similarity score is based on the Tanimoto method. The number of neighbours is defined as the number of compounds in the training set that have a distance equal to or smaller than 0.7 with respect to the test compound. The higher the number of neighbours, the more reliable the prediction for the test compound. Statistics of the distances are also calculated. Effectively no predictions are made for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated in Leadscope.

2. The Danish QSAR group

In addition to the general Leadscope structural applicability domain definition the Danish QSAR group has applied a further requirement to the applicability domain of the model. That is only positive predictions with a probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted. Predictions within the structural applicability domain but with probability between 0.5 to 0.7 or 0.3 to 0.5 are defined as positives out of applicability domain and negatives out of applicability domain, respectively. When these predictions are wed out the performance of the model in general increases at the expense of reduced model coverage.

### 5.2 Method used to assess the applicability domain

Leadscope does not generate predictions for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated.

Only positive predictions with probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted.

### 5.3 Software name and version for applicability domain assessment

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

### 5.4 Limits of applicability

The Danish QSAR group applies an overall definition of structures acceptable for QSAR processing which is applicable for all the in-house QSAR software, i.e. not only CASE Ultra. According to this definition accepted

structures are organic substances with an unambiguous structure, i.e. so-called discrete organics defined as: organic compounds with a defined two dimensional (2D) structure containing at least two carbon atoms, only certain atoms (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, and I), and not mixtures with two or more 'big components' when analyzed for ionic bonds (for a number of small known organic ions assumed not to affect toxicity the 'parent molecule' is accepted). Calculation 2D structures (SMILES and/or SDF) are generated by stripping off ions (of the accepted list given above). Thus, all the training set and prediction set chemicals are used in their non-ionized form. See 5.1 for further applicability domain definition.

6. Internal validation

6.1 Availability of the training set

Yes

6.2 Available information for the training set

CAS

SMILES

6.3 Data for each descriptor variable for the training set

No

6.4 Data for the dependent variable for the training set

All

6.5 Other information about the training set

836 compounds are in the training set: 418 positives and 418 negatives.

6.6 Pre-processing of data before modelling

The chemicals were classified in to mild or severe skin irritants based on the criteria described under 3.6.

Only structures acceptable for Leadscope were used in the final training set. That is only discrete organic chemicals as described in 5.4 were used. In case of replicate structures, one of the replicates was kept if all the compounds had the same activity and all were removed if they had different activity. No further structures accepted by the software were eliminated (i.e. outliers).

6.7 Statistics for goodness-of-fit

Not performed.

6.8 Robustness – Statistics obtained by leave-one-out cross-validation

Not performed. (It is not a preferred measurement for evaluating large models).

6.9 Robustness – Statistics obtained by leave-many-out cross-validation

A five times two-fold 50 % cross-validation was performed. This was done by randomly removing 50% of the full training set used to make the “mother model”, where the 50% contains the same ratio of positive and negatives as the full training set. A new model (validation submodel) was created on the remaining 50% using the same settings in Leadscope but with no information from the “mother model” regarding descriptor selection etc. The validation submodel was applied to predict the removed 50% (within the defined applicability domain for the submodel). Likewise, a validation submodel was made on the removed 50% of the training set and this model was used to predict the other 50% (within the defined applicability domain for this submodel). This procedure was repeated five times.

Predictions within the defined applicability domain of the ten validation submodels were pooled and Cooper’s statistics calculated. This gave the following results for the 65.1% (2723/(5*836)) of the predictions which were within the applicability domains of the respective submodels:

- Sensitivity (true positives / (true positives + false negatives)): 79.5%
- Specificity (true negatives / (true negatives + false positives)): 81.7%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 80.6%

6.10 Robustness - Statistics obtained by Y-scrambling

Not performed.

6.11 Robustness - Statistics obtained by bootstrap

Not performed.

6.12 Robustness - Statistics obtained by other methods

Not performed.

## 7. External validation

### 7.1 Availability of the external validation set

### 7.2 Available information for the external validation set

### 7.3 Data for each descriptor variable for the external validation set

### 7.4 Data for the dependent variable for the external validation set

### 7.5 Other information about the training 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

External validation not performed for this model.

## 8. Mechanistic interpretation

### 8.1 Mechanistic basis of the model

The global model identifies structural features and molecular descriptors which in the model development was found to be statistically significant associated with effect. Many predictions may indicate modes of action that are obvious for persons with expert knowledge for the endpoint.

### 8.2 A priori or posteriori mechanistic interpretation

A posteriori mechanistic interpretation. The identified structural features and molecular descriptors may provide basis for mechanistic interpretation.

### 8.3 Other information about the mechanistic interpretation

## 9. Miscellaneous information

### 9.1 Comments

The model can be applied to predict if a chemical causes severe or mild skin irritation *in vivo.*

### 9.2 Bibliography

EU classifications (1967) Directive 67/548/EEC- classification, packaging and labelling of dangerous substances. Annex I of Directive 67/548/EEC is available on: http://www.reach-compliance.eu/french/legislation/docs/launchers/launch-annex-1-67-548-EEC.html

CHEM-BANK (2000) Databanks of potentially hazardous chemicals: RTECS, OHMTADS, CHRIS, HSDB, IRIS, TSCA, NPG and ERG2000. USA. CHEM-BANK™, CD-ROM, SilverPlatter International N.V., November 2000.

OECD guideline 404 (2002) OECD Guidelines for the Testing of Chemicals No. 404: Acute Dermal Irritation/Corrosion. Organisation for Economic Cooperation and Development; Paris, France. Available online at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

RTECS (2012) Comprehensive guide to the RTECS. Registry of Toxic Effects of Chemical Substances 2012. Available on: http://ccinfoweb.ccohs.ca/rtecs/search.html

### 9.3 Supporting information