1. QSAR identifier

1.1. QSAR identifier (title):

   QSAR model for acute oral toxicity-in vitro method (IC50)

1.2. Other related models:

   Published in Toxicology Letters 170(2007) 214-222. (Correlation of the endpoint of the current QMRF - in vitro IC50 – with in vivo acute oral toxicity LD50, source of dataset)

1.3. Software coding the model:

   QSARModel 3.5.0 Molcode Ltd., Turu 2, Tartu, 51014, Estonia
   http://www.molcode.com

2. General information

2.1. Date of QMRF:

   30.08.2009

2.2. QMRF author(s) and contact details:

   [1] Indrek Tulp Molcode Ltd. Turu 2, Tartu, 51014, Estonia
   models@molcode.com http://www.molcode.com

   models@molcode.com http://www.molcode.com

   models@molcode.com http://www.molcode.com

   models@molcode.com http://www.molcode.com

   models@molcode.com http://www.molcode.com

   models@molcode.com http://www.molcode.com

   models@molcode.com http://www.molcode.com

   models@molcode.com http://www.molcode.com

   [9] Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia
   models@molcode.com http://www.molcode.com

   [10] Andres Kreegipuu Molcode Ltd. Turu 2, Tartu, 51014, Estonia
   models@molcode.com http://www.molcode.com

   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:
30.08.2009

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

2.8. Availability of information about the model:
All information in full detail is available.

2.9. Availability of another QMRF for exactly the same model:
No other QMRF available for the same model

3. Defining the endpoint - OECD Principle 1

3.1. Species:
Rat and mouse

3.2. Endpoint:
4. Human health effects
4.2. Acute oral toxicity

3.3. Comment on endpoint:
-

3.4. Endpoint units:
mmol/L

3.5. Dependent variable:
log(IC50) logarithm of the half maximal inhibitory concentration (IC50). The IC50 indicates how much of a particular substance (inhibitor) is needed to inhibit a given biological process (or component of a process) by half.

3.6. Experimental protocol:
The acute oral toxicity is determined using the OECD 420 and OECD 423 (EU B.1.bis. and B.1.tris.) test guidelines. Acute oral toxicity testing allows to obtain the information on the biologic/toxic activity of a chemical. Currently, the basis for toxicologic classification of chemicals is the median lethal dose (LD50, mg/kg b.w.), which is defined as the statistically derived dose required to kill half the members of a tested population. Animals are observed individually after dosing at least once during the first 30 minutes, periodically during the first 24 hours, with special attention given during the first 4 hours, and daily thereafter, for a total of 14 days. The European Chemicals regulation REACH (registration, evaluation, authorisation of chemicals) describes possible applications of
in vitro alternatives to replace in vivo testing. Demanding validation procedures are applied by European Centre for the Validation of Alternative Methods (ECVAM) to assure the suitability of any new in vitro method for regulatory purposes. This procedure has led to a number of internationally accepted in vitro alternatives for local endpoints (e.g. skin penetration and corrosion). However, in vitro alternatives for endpoints with complex physiology have shown important limitations proving difficulties to validate for a broad spectrum of chemicals. In fact, developed in vitro systems (biokinetic, cellular and molecular elements) can significantly reduce the number of animals needed for the toxicity testing of a broad range of compounds. There are 50 compounds in the dataset of the in vitro cytotoxicity (reference 1, median of several IC50 values of different experiments) consisting of drugs, agrochemicals and industrial chemicals. The IC50 values of tested substances were translated to logarithmic scale (logIC50) to reduce the range of the data. References 1. Freidig A.P., Dekkers S., Verwei M., Zvinavashe E., Bessems J.G.M., van de Sandt J.J.M., Development of a QSAR for worst case estimates of acute toxicity of chemically reactive compounds, Toxicology Letters 170 (2007) 214–222. 2. National Institute of Environmental Health Sciences National Institutes of Health US Public Health Service Department of Health and Human Services Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity Based on Recommendations from an International Workshop Organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), NIH Publication No. 01-4500, 2001.

3.7. Endpoint data quality and variability:
Experimental data from different sources has been validated as consistent (ref. Toxicology Letters 170 (2007) 214-222). Statistics: max value: 2.97 min value: -4.27 standard deviation: 1.90 skewness: -1.00

4. Defining the algorithm - OECD Principle 2
4.1. Type of model:
2D and 3D regression-based QSAR Multilinear regression model based on 3-D quantum chemical descriptors.
4.2. Explicit algorithm:
Multilinear regression QSAR model based on

\[ \text{Log(IC50)} = 5.48 - 41.64 \times \text{Global softness: } 1/(\text{LUMO} - \text{HOMO}) \text{ (AM1)} - 1.80 \times \text{Number of carbonyl groups} - 0.30 \times \text{Kier&Hall index (order 2)} \]

4.3. Descriptors in the model:
[2] Number of carbonyl groups
[3] Kier&Hall index (order 2)
4.4. Descriptor selection:
Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules (1-parameter equations: Fisher criterion and R2 over threshold (0.1), variance and t-test value over threshold (1.5), intercorrelation R2 with another descriptor not over threshold (0.3)), (2 parameter equations: intercorrelation squared coefficient below threshold (0.3), significant correlation with endpoint in terms of correlation coefficient and t-test (0.1, 1.5)). Stepwise trial of additional descriptors not significantly correlated to any already in the model.

4.5. Algorithm and descriptor generation:
1D, 2D, and 3D theoretical calculations quantum chemical descriptors derived from AM1 calculation. Model developed by using multilinear regression (least squares fit).

4.6. Software name and version for descriptor generation:
QSARModel 3.5.0
QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling
Molcode Ltd, Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

4.7. Descriptors/Chemicals ratio:
0.066

5.1. Description of the applicability domain of the model:
Applicability domain based on training set: By chemical identity: diverse set of chemically reactive organic compounds (alcohols, carboxylic acids, nitriles, aromatic compounds, sulfur and phosphorus compounds, etc) By descriptor value range: This model is suitable for compounds that have the descriptors in the following range: Global softness: 1/(LUMO - HOMO) (AM1)(min: 0.047, max: 0.169), Number of carbonyl groups (min: 0, max: 1), Kier&Hall index (order 2) (min: 0, max: 24.362)

5.2. Method used to assess the applicability domain:
presence of functional groups in structures Range of descriptor values in training set with ±30% confidence Descriptor values must fall between maximal and minimal descriptor values of training set ± 30%

5.3. Software name and version for applicability domain assessment:
QSARModel 3.5.0
QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling
Molcode Ltd, Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

5.4. Limits of applicability:
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: 45, negative: 14, positive values: 31

6.6. Pre-processing of data before modelling:
- 

6.7. Statistics for goodness-of-fit:
- \( R^2 = 0.85 \) (Correlation coefficient);
- \( s = 0.59 \) (Standard error of the estimate);
- \( F = 76.57 \) (Fisher function);

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:
- \( R^2_{cv} = 0.78 \) LOO

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:
- \( R^2_{cv} = 0.83 \) LMO

6.10. Robustness - Statistics obtained by Y-scrambling:
- 

6.11. Robustness - Statistics obtained by bootstrap:
- 

6.12. Robustness - Statistics obtained by other methods:
- ABC analysis (2:1 training : prediction) on sorted data divided into 3 subsets (A;B;C). Training set formed with 2/3 of the compounds (set A+B, A+C, B+C) and validation set consisted of 1/3 of the compounds (C, B, A) average \( R^2 \) (fitting) = 0.862 average \( R^2 \) (prediction) = 0.802

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:
data points: 5, negative: 2, positive values: 3

7.6. Experimental design of test set:
The full experimental dataset was sorted according to increasing values of logIC50 and each tenth compound was assigned to the test set.

7.7. Predictivity - Statistics obtained by external validation:
R2=0.802

7.8. Predictivity - Assessment of the external validation set:
The descriptors for the test set are in the limit of applicability

7.9. Comments on the external validation of the model:
The validation R2 for the test set is good.

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:
The acute oral toxicity depends on the stability and reactivity of the compound, the number of carbonyl groups and the shape of the molecule. Toxicity increases with increasing values of the descriptor Global softness: 1/(LUMO-HOMO)(AM1). The presence of the carbonyl group in the molecule accounts for a higher toxicity. The descriptor Kier&Hall index (order 2) shows that toxicity is also influenced by the shape and size of the molecule.

8.2. A priori or a posteriori mechanistic interpretation:
a posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments

8.3. Other Information about the mechanistic Interpretation:
The descriptor Global softness: 1/(LUMO - HOMO) (AM1)[1/eV] gives information about the reactivity and stability of the molecule. Increased value for Global softness indicates a higher reactivity, so toxicity increases for more reactive compounds. The descriptor Number of carbonyl groups accounts for the presence of carbonyl groups in the molecule. The Kier&Hall index (order 2) gives information about different aspects of atom connectivity within a molecule, about the branching of the ring structures and the flexibility. The proposed mechanism based on the model agrees well with literature: Toxicology Letters 170(2007) 214-222.

9. Miscellaneous information

9.1. Comments:
Dataset contains rat or mouse intravenous and oral in vivo IC50 values. So far, the most promising approaches for predicting acute oral toxicity in mammals is by in vitro–in vivo correlations using cytotoxicity data (that's why the additional set contains cytotoxicity data for the same
<table>
<thead>
<tr>
<th>Training set(s)</th>
<th>Test set(s)</th>
<th>Supporting information</th>
</tr>
</thead>
</table>

**Summary (ECB Inventory)**

10.1. QMRF number:
10.2. Publication date:
10.3. Keywords:

10.4. Comments: