Model QSAR2 for thyroperoxidase (TPO) inhibition *in vitro* (U.S. EPA NCCT data)

1. QSAR identifier

1.1 QSAR identifier (title)

Leadscope Enterprise model QSAR2 for thyroperoxidase (TPO) inhibition *in vitro* (U.S. EPA NCCT data), Danish QSAR Group at DTU Food.

1.2 Other related models

Leadscope Enterprise model QSAR1 for thyroperoxidase (TPO) inhibition *in vitro* (U.S. EPA NCCT data), Danish QSAR Group at DTU Food.

1.3. Software coding the model

Leadscope Predictive Data Miner, a component of Leadscope Enterprise, version 3.5.

2. General information

2.1 Date of QMRF

June 2018.

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;

Kyrylo Klimenko;

National Food Institute at the Technical University of Denmark;

2.3 Date of QMRF update(s)

None

2.4 QMRF update(s)

None

2.5 Model developer(s) and contact details

Sine Rosenberg

PhD student 2013-2017 at the National Food Institute at the Technical University of Denmark

Eva Wedebye

National Food Institute at the Technical University of Denmark

Nikolai Nikolov

National Food Institute at the Technical University of Denmark

Marianne Dybdahl

National Food Institute at the Technical University of Denmark

2.6 Date of model development and/or publication

August 2017.

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 data set is non-proprietary and was kindly provided by U.S. EPA NCCT with chemical structure information and HTS screening results for TPO inhibition. 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

Thyroid tissue from male Long Evans rats

3.2 Endpoint

QMRF 4. Human Health Effects

QMRF 4.18.c. Endocrine Activity. Other (rat thyroperoxidase enzyme inhibition)

3.3 Comment on endpoint

The assay applied for the training set data is described in Paul et al. 2014 and is based on Rat thyroid microsomes and a fluorescent peroxidase substrate (Amplex UltraRed, AUR).

Taken from Paul et al. 2014: *Toxicants that inhibit TPO activity prevent iodothyronine production in the thyroid gland. TPO is a heme-containing, multifunction enzyme critical to thyroid hormone (TH) synthesis located at the apical membrane of follicular thyroid cells. TPO catalyzes the oxidation of iodide to hypoiodate, the addition of hypoiodate to tyrosyl residues on thyroglobulin (Tg), and concurrent oxidative coupling of iodinated tyrosyl residues to form iodothyronine hormones triiodothyronine (T3) and thyroxine (T4).*

In animal models and humans altered cognition, socialization and motor function as well as hearing loss have been observed following moderate to severe hypothyroidism. Even low levels of TH insufficiency during fetal development may result in measurable IQ deficits in children. In adulthood, dysregulated TH levels can give reversible clinical symptoms of hypo- or hyperthyroidism and are correlated with pathological processes involved in adverse outcomes such as cancer, obesity and type II diabetes mellitus.

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Inhibition of the TPO in vitro, positive or negative.

3.6 Experimental protocol

The experimental protocol is described in detail in Paul et al. 2014. Briefly, the assay measures the fluorescence intensity from the commercial peroxidase substrate, Amplex®UltraRed (AUR), which is converted to Amplex UltroxRed by a peroxidase in the presence of hydrogen peroxide. A decrease in fluorescence intensity in response to a chemical is an indirect measure of TPO inhibition.

All chemicals were initially screened at a single, high concentration (~87.5µM). The chemicals associated with 20% or greater decreases in maximal TPO activity were subsequently screened for possible concentration-response. The concentration-response data were processed using the ToxCast data pipeline whereby each chemical was assigned a 'hit-call' of 1 if active in AUR-TPO, or a 'hit-call' of 0 if inactive in AUR-TPO. Actives in the AUR-TPO assay were further processed through a selectivity filtering algorithm, which integrates results from cytotoxicity and luciferase inhibition assays to identify possible non-specific positive results in the AUR-TPO assay.

We classified the chemicals into three categories: 1) chemicals that had a <20% activity decrease in the single, high concentration screening or had been assigned a 'hit-call' of 0 in the concentration-response AUR-TPO screening were classified as inactive in this assay; 2) chemicals with a 'hit-call' of 1 in AUR-TPO and a selectivity score greater than 1 were classified as active for TPO inhibition; and 3) chemicals with a 'hit-call' of 1 in AUR-TPO but with a selectivity score of 1 or less were classified as inconclusive for TPO inhibition and not used in the training set of the model.

3.7 Endpoint data quality and variability

The datasets originate from the same source, i.e. U.S. EPA NCCT. All chemicals have been screened in the same testing protocols and undergone the same data processing, and this has likely contributed to decrease the experimental variability. The quality of the AUR-TPO assay has been assessed previously which indicated excellent performance with robust Z-prime factor from 0.77 to 0.83, where values above 0.5 generally indicate excellent performance to distinguish between actives and inactives, and high intralaboratory repeatability with the robust coefficient of variance being 3–4%. However, no measures of the reproducibility of the overall positive and negative end calls as used for QSAR model was available. Still, the data in training set 1 and 2 and the test set were assessed to be of high quality and expected to be a good basis for QSAR model development.

4. Defining the algorithm

4.1 Type of model

A categorical QSAR model based on structural features and numeric molecular descriptors.

4.2 Explicit algorithm

This is a categorical QSAR model made by use of partial logistic regression (PLR). The model is a 'cocktail' composite model that integrates a so-called single model, see 4.4, and a Leadscope composite model with 6 sub-models, see 4.4, i.e. the cocktail composite model contains 7 sub-models. 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,

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)), with subsequent removal of redundant structural features, as well as the calculated molecular descriptors in the descriptor selection.

Based on model performance as measured by a Leadscope Predictive Data Miner cross-validation the model developed using approach integrating number 2 and 5 Into a cocktail composite model was chosen.

For this model scaffolds were generated by Leadscope for the training set structures and added to the Leadscope library of structural features. The number of structural features was then reduced further using the built-in filter to remove similar (reduntant) features (the "less similar" features removed). Descriptors were then automatically selected among the remaining 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 sub-models and they may therefore have a tendency to give overoptimistic validation results.

In this model because of the categorical outcome in the response variable PLR was used to build the predictive model. Because of the unbalanced training set (i.e. 130 positives and 747 negatives) 7 sub-models for smaller individual training sets were made. The descriptors for each of the sub-models were automatically selected from the Leadscope feature library based solely on the training set compounds used to build the individual sub-model and was not affected by the full training set chemicals . Therefore, a different number of descriptors (structural features and molecular descriptors) were selected and distributed on varying number of PLS factors for each sub-model.

4.6 Software name and version for descriptor generation

Leadscope Predictive Data Miner, a component of Leadscope Enterprise, version 3.5.

4.7 Descriptors/chemicals ratio

As this model is a composite model consisting of 7 sub-models with varying training set size and using different descriptors and number of PLS factors (see 4.5), an overall descriptor/chemical ratio for this model cannot be calculated. The data for individual models, as well as, mean, minimum and maximum descriptors/chemicals ratios are as follows:

Name of the model	Chemicals	Descriptors	PLS factors
TC_TPO_Multiple_Scaffolds_Reduced_Model-1	1519	263	4
TC_TPO_Multiple_Scaffolds_Reduced_Model-2	460	105	2
TC_TPO_Multiple_Scaffolds_Reduced_Model-3	460	175	3
TC_TPO_Multiple_Scaffolds_Reduced_Model-4	460	167	3
TC_TPO_Multiple_Scaffolds_Reduced_Model-5	460	160	1
TC_TPO_Multiple_Scaffolds_Reduced_Model-6	460	173	3
TC_TPO_Single_Scaffolds_Reduced_Model	460	169	5

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 weeded 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 3.5.

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 Leadscope. 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, downloadable at http://qsar.food.dtu.dk/download/TPO_inhibition_QSAR_training_set

6.2 Available information for the training set

The zip archive has an xls file with “Presence in test set marked with *” column marking structures that must be removed to obtain the QSAR1 model training set. Besides the TPO inhibition calls, the file contains CAS and SMILES for each substance.

6.3 Data for each descriptor variable for the training set

No

6.4 Data for the dependent variable for the training set

All. Column “TC TPO inhibition activity” in the Excel file from the downloadable zip archive indicates if the compound is Active (positive) or Inactive (negative).

6.5 Other information about the training set

1519 compounds are in the training set: 230 positives and 1289 negatives. The training set was created by combining the training set and the external validation set for QSAR1 for TPO inhibition, see Rosenberg et al. 2017 for more information.

6.6 Pre-processing of data before modelling

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

These results were obtained using Leadscope Predictive Data Miner v.3.2.4 and reported in Rosenberg et al. 2017, the current model was re-modelled using Leadscope Predictive Data Miner v.3.5 and is expected to be associated with similar accuracy:

A ten times two-fold 50 % cross-validation was performed (not using the Leadscope Predictive Data Miner built-in procedure, but by a DTU procedure, see Rosenberg et al. 2017). 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 sub-model) 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 sub-model was applied to predict the removed 50% (within the defined applicability domain for the sub-model). Likewise, a validation sub-model 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 sub-model). This procedure was repeated ten times.

Predictions within the defined applicability domain of the twenty validation sub-models were pooled and Cooper's statistics calculated. This gave the following results for the predictions which were within the applicability domains of the respective sub-models:

- Sensitivity (true positives / (true positives + false negatives)): 75.6% (SD 5.0)
- Specificity (true negatives / (true negatives + false positives)): 89.8% (SD 1.5)
- Balanced Accuracy ((Sensitivity + Specificity) /2): 82.7% (SD 2.2)
- Coverage ((In-Domain predictions) / (All predictions): 57.8% (SD 5.4)

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

Not performed

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 validation 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

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

None

9.2 Bibliography

Rosenberg , S.A., Watt, E.D., Judson, R.S., Simmons, S.O., Friedman, K. Paul, Dybdahl, M., Nikolov, N.G., Wedebye, E.B. (2017) QSAR models for thyroperoxidase inhibition and screening of U.S. and EU chemical inventories. *Computational toxicology*, 4, 11-21.

Paul K.B., Hedge J.M., Rotroff D.M., Hornung M.W., Crofton K.M., Simmons S.O., Development of a Thyroperoxidase Inhibition Assay for High-Throughput Screening, Chem. Res. Toxicol. 27 (2014) 387–399.