(Q)SAR Model Reporting Format (QMRF)

(The present QMRF v.2.1 is prepared in accordance with (Q)SAR Assessment Framework (QAF) document developed by OECD)

(https://one.oecd.org/document/ENV/CBC/MONO(2023)32/ANN1/en/pdf**)**

Welcome

Model version: Skin sensitization with autoxidation v.26.32
Platform version: OASIS TIMES v.2.33.1
Name: Skin sensitization with autoxidation
Author: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Date: 15 April, 2024
e-mail: omekenya@btu.bg
www: http://www.oasis-lmc.org/

Section 1. QSAR identifier

1.1. QSAR identifier (title)

Skin sensitization with autoxidation v.26.32

1.2. Other related models

Not applicable

1.3. Software coding the model

Model version: Skin sensitization with autoxidation v.26.32
Platform version: OASIS TIMES v.2.33.1
Name: Skin sensitization with autoxidation
Developer: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Coding language: Delphi 10.2

Section 2. General information

2.0. Abstract

The current Skin Sensitization (SS) model was developed using a dataset of 1405 chemicals (39 of which have proprietary skin sensitization data) tested by Local Lymph Node Assay (LLNA) or Guinea Pig Maximization Test (GPMT). A unifying scale was derived through evaluating the correlation and concordance of those chemicals that existed in two datasets. The skin sensitization model integrates a simulator of skin metabolism together with a list of alerts for protein binding built based on the training set chemicals and expert knowledge.

The model was implemented in TIMES system that allows in the same platform to be predicted metabolism of chemicals and toxicity resulting from their metabolic activation. Autoxidation (AU) of chemical is also accounted for.

2.1. Date of QMRF

15 April 2024

2.2. QMRF author(s) and contact details

Name: Laboratory of Mathematical Chemistry
Affiliation: Laboratory of Mathematical Chemistry, University "Prof. As. Zlatarov", "Yakimov" St. #1, 8010 Bourgas, BULGARIA
URL: http://www.oasis-lmc.org
E-mail: omekenya@btu.bg

2.3. Date of QMRF update(s)

November 2013; June 2014; January 2015; June 2015; January 2016; June 2016; June 2017; July 2018; January 2020; January 2022; March 2023; April 2024

2.4. QMRF update(s)

Information which has been modified:
Sections 1.1 QSAR identifier (title)**; Sections 1.3** Software coding the model; **Section 2.** General information; **Sections 2.0** Abstract; **Sections 2.1** Date of QMRF; **Sections 2.3** Date of QMRF update(s); **Sections 2.8** Availability of information about the model; **Section 3.2** Endpoint; **Sections 3.3** Comment on endpoint; **Section 4.6** Software name and version for descriptor generation

2.5. Model developer(s) and contact details

The TIMES-SS model was developed by the Laboratory of Mathematical Chemistry using funding and data from a Consortium comprising Industry (ExxonMobil, P&G, Unilever, L'Oreal, Dow Chemicals, DuPont, Givaudan and RIFM) and a Regulatory Agency (DK-EPA).

S. Dimitrov, L. Low, G. Patlewicz, P. Kern, G. Dimitrova, M. Comber, R. Philips, J. Niemela, Bailey, O. Mekenyan

Affiliation: Laboratory of Mathematical Chemistry, University "Prof. As. Zlatarov", "Yakimov" St. #1, 8010 Bourgas, BULGARIA
URL: http://www.oasis-lmc.org
E-mail: omekenya@btu.bg

2.6. Date of model development and/or publication

Date of the model development: 2005 April

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

1. S. Dimitrov, L. Low, G. Patlewicz, P. Kern, G. Dimitrova, M. Comber, R. Philips, J. Niemela, P. Bailey, O. Mekenyan. Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. *International Journal of Toxicology*, 24:189-204, (2005).
2. O. Mekenyan, S. Dimitrov, T. Pavlov, G. Dimitrova, M. Todorov, P. Petkov & S. Kotov. 2012. Simulation of chemical metabolism for fate and hazard assessment. V. Mammalian hazard assessment, *SAR and QSAR in Environmental Research*, Vol. 23, 553-606.

2.8. Availability of information about the model

Skin Sensitization (SS) model is proprietary and its use is subject of licence agreement.
Information that cannot be disclosed:

- External validation sets,
- Proprietary chemicals,
- Source code.

For more details, please contact Professor Ovanes Mekenyan: omekenya@btu.bg
Details of the model are provided in the sections bellow as well as in the following link:
http://oasis-lmc.org/products/models/human-health-endpoints/skin-sensitization.aspx

2.9. Availability of another QMRF for exactly the same model

Not available.

Section 3. Defining the endpoint – OECD Principle 1

3.1. Species

mouse – used in Local Lymph Node Assay;
guinea pigs – used in Guinea Pig Maximization Test

3.2. Endpoint

In vivo Skin sensitization according to:

- OECD TG 429 https://www.oecd-ilibrary.org/environment/test-no-429-skin-sensitisation_9789264071100-en
- OECD TG 406 https://www.oecd-ilibrary.org/environment/test-no-406-skin-sensitisation_9789264070660-en

3.3. Comment on endpoint

Skin sensitization resulting in allergic contact dermatitis is a common occupational and environmental health issue. Current approaches to concretely assess skin sensitization potential, is carried out through *in vivo* testing: Local Lymph Node Assay (LLNA) and Guinea Pig Maximization Test (GPMT). The training set of TIMES Skin sensitization model contains only chemicals having:

- LLNA EC3 values generated by test protocols equivalent or similar to OECD TG 429 or
- GPMT data generated by test protocols equivalent or similar to OECD TG 406.

3.4. Endpoint units

LLNA – EC3, %

GPMT - % of animals showing reaction of skin

3.5. Dependent variable

Obs. Skin Sensitization effect

3.6. Experimental protocol

OECD technical guidelines 429 and 406 describe the LLNA (the murine local lymph node assay) and GPMT/Buehler tests (the guinea pig maximization test) that are available.

3.7. Endpoint data quality and variability

High quality. The model was derived from a data set compiled from chemicals tested in the LLNA or GPMT. All data was QA-ed and generated by test protocols equivalent or similar to OECD TG 429 and 406. Detailed information is provided in **Appendix 1**.

Section 4. Defining the algorithm – OECD Principle 2

4.1. Type of model

Structural alerts based model

4.2. Explicit algorithm

TIMES-SS model aims to encode structure toxicity and structure metabolism relationships through a number of transformations simulating skin metabolism and interaction of the generated reactive metabolites with skin proteins. The skin metabolism simulator mimics metabolism using 2D structural information. The autoxidation (abiotic oxidation) of chemicals is also accounted for. A training set of diverse chemicals was compiled and their skin sensitization potency assigned to one of three classes. These three classes were Strong, Weak or Non sensitizing.

4.3. Descriptors in the model

Descriptors in the model are structural alerts related to interactions with skin Proteins. Each alert is supported by detailed mechanistic description of the interaction with proteins provided as additional explanatory textual information. To assess the reactivity of some specific alerts addition requirements for logKow were set in the model.

4.4. Descriptor section

Table 1 summarizes the main characteristics of each Protein binding alerts in the TIMES Skin sensitization model:

- Alert name (corresponding to the name of the chemical class which is addressed);
- Performance of alert (correct/incorrect predictions) which is estimated based on proportion of observed positive chemicals from all chemicals captured by the alert. Performance of each alert is provided with its confidence range. As smaller is the size of local training sets as wider are the confidence ranges and vice versa.
- P-values addressing the reliability of alert performance estimation and taking into account possible bias of positive/negative chemicals in the training set of the model. Low p-values could be obtained only if both are satisfied:
    - The number of chemicals in local training set is high enough;
    - The alert performance is significantly higher than the proportion of positive/negative chemicals in the model training set, i.e. so-called naïve alert.

Analogically, high p-values could be obtained in case of:

- Small number of local training set chemicals (1-2 chemicals); or
- Performance comparable to the performance of the naïve alert.

High performance associated with low *p-values* indicate for High Reliability of alerts.

The above statistical measures along with the underlying mathematical formalisms are discussed in details in **Section 6** (Internal validation).

Table 1. Main characteristics of the Protein binding alerts in the TIMES Skin sensitization model.

No	Alert	Corr	Incorr	Performance $_{est.}$ [1)]	*p-value*[1)]
1	1,2-Dicarbonyls, 1,3-Dicarbonyls	17	0	0.947 (0.847 ÷ 1.000)	2.2E-7
2	Activated Aryl and Heteroaryl Compounds	31	2	0.914 (0.823 ÷ 0.990)	1.7E-10
3	Azlactones and Unsaturated Lactone Derivatives	9	0	0.909 (0.741 ÷ 1.000)	0.0003
4	Activated (Di)Aryl Esters	15	1	0.889 (0.748 ÷ 0.997)	1.3E-5
5	Sulfonates	6	0	0.875 (0.652 ÷ 1.000)	0.0043
6	Thiols and Disulfide Compounds	19	2	0.870 (0.735 ÷ 0.983)	2.9E-6
7	Alkyl Halides	25	3	0.867 (0.746 ÷ 0.972)	1.3E-7
8	Anhydrides (Sulphur Analogues of Anhydrides)	5	0	0.857 (0.607 ÷ 1.000)	0.010
9	Bis-Epoxides	5	0	0.857 (0.607 ÷ 1.000)	0.010
10	Nitrosoalkenes	5	0	0.857 (0.607 ÷ 1.000)	0.010
11	alpha, beta-Carbonyl Compounds with Polarized Triple Bond	4	0	0.833 (0.549 ÷ 1.000)	0.026
12	alpha, beta-Unsaturated Oximes	4	0	0.833 (0.549 ÷ 1.000)	0.026
13	Conjugated Systems with Electron Withdrawing Groups	4	0	0.833 (0.549 ÷ 1.000)	0.026
14	Dithioesters	4	0	0.833 (0.549 ÷ 1.000)	0.026
15	Isothiazolone Derivatives	4	0	0.833 (0.549 ÷ 1.000)	0.026
16	Isothiocyanates and Isocyanates	4	0	0.833 (0.549 ÷ 1.000)	0.026
17	Generated Free Radicals/Ion Radicals	16	3	0.810 (0.645 ÷ 0.957)	0.0001
18	Bis Aldehydes	7	1	0.800 (0.567 ÷ 0.991)	0.0090
19	Activated Alkyl Diesters	3	0	0.800 (0.473 ÷ 1.000)	0.065
20	Lactones	3	0	0.800 (0.473 ÷ 1.000)	0.065
21	N-Nitroso Compounds	3	0	0.800 (0.473 ÷ 1.000)	0.065
22	Epoxides, Aziridines and Sulfuranes	22	5	0.793 (0.647 ÷ 0.928)	1.7E-5
23	alpha, beta-Carbonyl Compounds with Polarized Double Bonds	48	13	0.778 (0.675 ÷ 0.876)	2.2E-9

24	(Thio)Acyl and (Thio)Carbamoyl Halides, Cyanides, Azides, etc.	6	1	0.778 (0.524 ÷ 0.989)	0.019
25	Di-Substituted alpha,beta-Unsaturated Aldehydes	6	1	0.778 (0.524 ÷ 0.989)	0.019
26	alpha, beta-Aldehydes	16	4	0.773 (0.601 ÷ 0.931)	0.0004
27	Amides	9	2	0.769 (0.550 ÷ 0.964)	0.0064
28	Quinone Methide(s)/Imines, Quinoide Oxime Structure, Nitroquinones, Naphthoquinone(s)/Imines	83	25	0.764 (0.684 ÷ 0.841)	< 1.0E-10
29	Aldehydes	58	18	0.756 (0.661 ÷ 0.848)	4.8E-10
30	alpha-Ketoesters	2	0	0.750 (0.368 ÷ 1.000)	0.16
31	Benzoyl Phosphine Oxides	2	0	0.750 (0.368 ÷ 1.000)	0.16
32	Benzyl or Phenethyl Salicylates	2	0	0.750 (0.368 ÷ 1.000)	0.16
33	N-alkyl thio succinimides	2	0	0.750 (0.368 ÷ 1.000)	0.16
34	Organic peroxides	2	0	0.750 (0.368 ÷ 1.000)	0.16
35	Phenyl Carbonates	2	0	0.750 (0.368 ÷ 1.000)	0.16
36	Sulfates	2	0	0.750 (0.368 ÷ 1.000)	0.16
37	Sulphonyl Halides or Cyanides	2	0	0.750 (0.368 ÷ 1.000)	0.16
38	Thiocyanates	2	0	0.750 (0.368 ÷ 1.000)	0.16
39	Thiourea Compounds	2	0	0.750 (0.368 ÷ 1.000)	0.16
40	alpha-Activated Haloalkanes	9	3	0.714 (0.486 ÷ 0.926)	0.016
41	alpha-Activated Benzyls	4	1	0.714 (0.409 ÷ 0.982)	0.089
42	Bifunctional alpha, beta-Carbonyl Containing Compounds	5	2	0.667 (0.379 ÷ 0.935)	0.099
43	Active Cyclic Agents	3	1	0.667 (0.330 ÷ 0.974)	0.18
44	Activated (Thio)Esters	1	0	0.667 (0.224 ÷ 1.000)	0.40
45	alpha, beta-Carbonyl Compounds with Polarized Double Bonds - Methacrylate type	1	0	0.667 (0.224 ÷ 1.000)	0.40
46	Anthraquinones	1	0	0.667 (0.224 ÷ 1.000)	0.40
47	Azomethynes with a Sulfo Leaving Group	1	0	0.667 (0.224 ÷ 1.000)	0.40
48	Carbenium Ion	1	0	0.667 (0.224 ÷ 1.000)	0.40
49	Carbodiimides	1	0	0.667 (0.224 ÷ 1.000)	0.40
50	Diacyl Peroxides, Anhydrides (Sulphur Analogues of Diacyl Peroxides)	1	0	0.667 (0.224 ÷ 1.000)	0.40
51	Dithiocarbamates	1	0	0.667 (0.224 ÷ 1.000)	0.40
52	Guanidines	1	0	0.667 (0.224 ÷ 1.000)	0.40
53	Iodoalkynes	1	0	0.667 (0.224 ÷ 1.000)	0.40
54	Isothiazolidin-3-ones (Sulphur) and Isothiazolone Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.40

55	N-Sulfonylazomethynes	1	0	0.667 (0.224 ÷ 1.000)	0.40
56	Phenolic esters	1	0	0.667 (0.224 ÷ 1.000)	0.40
57	Phosphite esters	1	0	0.667 (0.224 ÷ 1.000)	0.40
58	Polarised Alkene - Alkenyl Pyridines, Pyrazines, Pyrimidines or Triazines	1	0	0.667 (0.224 ÷ 1.000)	0.40
59	Polarised Alkenes - Sulfones	1	0	0.667 (0.224 ÷ 1.000)	0.40
60	Pyrazolones and Pyrazolidinones	1	0	0.667 (0.224 ÷ 1.000)	0.40
61	Vinyl-Type Compounds with Electron Withdrawing Groups	1	0	0.667 (0.224 ÷ 1.000)	0.40
62	(Thio) Phosphates	2	1	0.600 (0.228 ÷ 0.956)	0.35
63	C-Nitroso Compounds	2	1	0.600 (0.228 ÷ 0.956)	0.35
64	Activated Carbonyl Compounds	6	4	0.583 (0.318 ÷ 0.841)	0.17
65	Activated Alkyl Esters and Thioesters	4	3	0.556 (0.254 ÷ 0.851)	0.29
66	Hydroperoxides	45	42	0.517 (0.414 ÷ 0.620)	0.022
67	Carbamates	2	2	0.500 (0.147 ÷ 0.853)	0.52
68	beta-Lactams	1	1	0.500 (0.094 ÷ 0.906)	0.64
69	Cyanoalkenes	1	1	0.500 (0.094 ÷ 0.906)	0.64
70	Ketones	17	39	0.310 (0.195 ÷ 0.429)	0.94

[1)] Confidence ranges and p-value are calculated at 95% confidence level

Alerts which are not supported by chemicals from the training set (theoretical alerts) are not included in Table 1.

4.5. Algorithm and descriptor generation

For derivation of each alert mechanistically justifiable structural fragments for interaction with skin proteins are identified from the chemicals having positive data in the training set and/or suggested by experts.

4.6. Software name and version for descriptor generation

TIMES Skin sensitization model v.26.32

4.7. Chemicals/Descriptors ratio

Provided in Section 4.4.

Section 5. Defining the applicability domain of the model – OECD Principle 3

5.1. Description of the applicability domain of the model

The domain consists of the following sub-domain layers:

1. General parametric requirements.

   The variations of molecular parameters that may affect the quality of the measured endpoint significantly are included here (such as molecular weight, etc.). The domain of general parametric includes the range of variation of hydrophobicity (log KOW) and Molecular weight (MW) of chemicals in training set.

2. Structural domain.

The structural component of the model is based on the structural similarity between chemicals in the training set which were correctly predicted by the model. The structural neighborhood of atom-centered fragments (accounting for the first neighbours) extracted from correctly and incorrectly predicted parent structures from the training set is used to determine this similarity.

The target chemical could contain the following types of ACF:

- Fragments present in correctly predicted training chemicals only (i.e. correct fragments),
- Fragments found both in correctly and non-correctly predicted training chemicals (i.e. fuzzy fragments). These fragments are treated as correct fragments,
- Fragments present in non-correctly predicted training chemicals only (i.e. incorrect fragments),
- Fragments not present in the training chemicals (i.e. unknown fragments).

A chemical belongs to the structural domain of the model if it could be partitioned only on correct fragments. The user is able to analyse how important are unknown and incorrect fragments (if present in the target) and to make a decision about their effect on the quality of prediction. The distribution of structural characteristics of the target chemical and accepted thresholds is used as a criterion to determine how well the target is represented in the structural space of correctly predicted chemicals. The accepted domain thresholds for Skin sensitization model are as follows:

- Correct = 100%
- Incorrect = 0%

A chemical is considered In Domain if it is classified to belong to all sub-domain levels. The information implemented in the applicability domain is extracted from the correctly predicted training chemicals used to build the model and in this respect the applicability domain determines practically the interpolation space of the model.

5.2. Method used to assess the applicability domain

The approach used to determine and assess the domain is described in:

Dimitrov S, Dimitrova G., Pavlov T., Dimitrova N., Patlewicz G., Niemela J., Mekenyan O., A stepwise approach for defining the applicability domain of SAR and QSAR models, *J. Chem. Inf. Model.*, 45, 839-849 (2005).

5.3. Software name and version for the applicability domain assessment

The LMC software OASIS Domain Manager v.1.13 (which is embedded in OASIS platform) is used to determine the applicability domain.
http://oasis-lmc.org/products/software/domain-manager.aspx

5.4. Limits of applicability

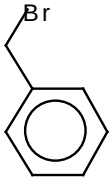
Applicability domain of the Skin sensitization model includes the following sub-domain layers: general parametric requirements, structural features and alerts reliability.

- General properties requirements:

As described in the Section 5.1.1, parametric domain of the model is derived based on Log *Kow* and *MW*. Example demonstrating belonging of a training set chemical to the parametric layer of the model domain is provided below:

Example chemical:

- CAS: 100-39-0
- Name: Benzyl bromide
- 2D Depiction:



Property	Domain	Example chemical
log K_{OW}	[-13.7; 33.5]	2.88
MW, Da	[6.9; 1353.4]	171.03

* K_{OW} is calculated by EPIWin

The values of *logKow* and *MW* of the example chemical are within the ranges of these parameters extracted from the whole training set of the model. Hence, with respect to the general parametric requirements, the target chemical is estimated to be *In Domian*.

- Structural features

Structural domain of the model is extracted from 1405 training chemicals containing:

- 3 469 correct fragments,
- 652 fuzzy fragments (treated as correct fragments),
- 598 incorrect fragments.

- Alerts reliability

Reliability of alerts is estimated based on:

- Alert Performance (Concordance) within the local training set chemicals (C);
- Number of the local training sets (N);
- Mechanistic justification (M).

According to these criteria, there are four reliability estimates for the alerts in the models:

- High reliability alerts ($C \geq 0.6$, $N \geq 5$, M)
- Low reliability alerts ($C \leq 0.6$, $N \geq 5$, M)
- Undetermined alerts ($1 < N < 5$, M)
- Undetermined theoretical alerts (M).

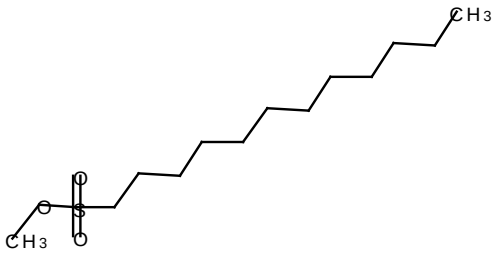
Example of a chemical belonging to alert with “High reliability”:

Chemical ID:

CAS: 2374-65-4

Name: 4-Methyl dodecane sulphonate

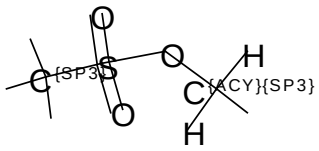
Depiction:



Belonging to alert:

Name: Sulfonates

Structural boundaries:



Reliability:

"High reliability" based on C=1; N=6 and M.

Currently, information for alerts reliability is provided in the model reports.

Section 6. Defining goodness-of-fit and robustness (internal validation) – OECD Principle 4

6.1. Availability of the training set

The training set consists of 1405 chemicals, including, 565 positive and 840 negative.

6.2. Available information for the training set

Chemical names, CAS numbers, SMILES and data source information are available.

6.3. Data for each descriptor variable for the training set

Descriptors in the models are structural alerts. The main characteristics of each alert are provided in Table 1 (Section 4.4).

6.4. Data for the dependent variable for the training set

The training set of 1405 chemicals includes:

- 565 chemicals with positive data for skin sensitization
- 840 chemicals with negative data from skin sensitization test.

Distribution of positive/negative chemicals in the training set of model is used for estimating performance and confidence range of the so-called *naïve alert* which is 0.402 (0.377 ÷ 0.428) [1)].

[1)] Confidence range is calculated at 95% confidence level

6.5. Other information about the training set

The current skin sensitization model was developed using a dataset of 1405 chemicals (39 of which have proprietary data for skin sensitization) tested by Local Lymph Node Assay

(LLNA) or Guinea Pig Maximization Test (GPMT).

A unifying scale was derived evaluating the correlation and concordance of those chemicals that exist in two datasets:

Unified skin sensitization scale	LLNA	GPMT
Strong	Extreme, Strong & Moderate	Strong & Moderate
Weak	Weak	Weak
Non	Non	Non

The distribution of training set chemicals having skin sensitization experimental data among the sensitization classes is as follows:

- 370 are Strong skin sensitizers
- 195 are Weak skin sensitizers
- 840 are Non skin sensitizers

6.6. Pre-processing of data before modelling

Not available

6.7. Statistics for goodness-of-fit

During the internal validation the original training set is separated many times randomly into two parts – one becomes a training set and the other becomes a test set. The model is re-derived many times using each new training set. Then, performance is estimated for the training sets and test sets. The averaged value of all training set performances is compared to the averaged value of all test set performances in order to assess the amount of optimism in the goodness-of-fit (GOF optimism) in the original model. GOF optimism is calculated as average performance over training sets minus average performance over test sets. Results are provided in Table 2.

Table 2. Performance of the original model over its training set (goodness-of-fit, GOF) vs. expected performance over set different from the training set (GOF – GOF optimism).

	Performance $_{est.}$ [1)]	***p-value*** [1)]	**External performance** [2) 3)]
All predictions (accuracy)	0.846 (0.828 ÷ 0.865)	< 1.0E-10	0.828
Positive chemicals (sensitivity)	0.878 (0.851 ÷ 0.905)	< 1.0E-10	0.828
Negative chemicals (specificity)	0.824 (0.798 ÷ 0.850)	< 1.0E-10	0.818

1) Confidence ranges and *p-value* are calculated at 95% confidence level.

2) Estimated performance for training set minus GOF optimism calculated from internal validation.

[3)] Estimation of expected performance over sets different from training sets.

Addition information including mathematical formalism underlying the above statistical measures is provided in **Appendix 2**.

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

Not performed

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

Method 1. *k-fold* cross-validation

In ***k-fold cross-validation*** the original training set is partitioned into *k* equally sized subsets. Each time a single subset is used as a test set and the remaining *k-1* subsets are used as training set. In this manner the process is repeated *k* times and each data from the original training set is used once as a test data and *k-1* times as a training data. The advantage of this method is that any data is used for both training and validation and each data is used exactly once as a test data. Commonly the 10-fold cross-validation is used (90% training data, 10% test data). In addition, 4-fold cross validation (75% training data, 25% test data) is also performed and the results from both procedures are provided in Table 3.

Table 3. Results from *k-fold* (*10-fold* and *4-fold*) cross-validation.

	10-fold		**4-fold**	
	Training sets	**Test sets**		**Training sets**
Unique chemicals, %	90.0 (89.9 ÷ 90.1)	10.0 (9.9 ÷ 10.1)	Unique chemicals, %	90.0 (89.9 ÷ 90.1)
Performance $_{est.}$, all predictions (accuracy)	0.846 (0.834 ÷ 0.859)	0.824 (0.694 ÷ 0.955)	Performance $_{est.}$, all predictions (accuracy)	0.846 (0.834 ÷ 0.859)
p-value, accuracy	< 1.0E-10	9.8E-10	*p-value*, accuracy	< 1.0E-10
Performance $_{est.}$, positive chemicals (sensitivity)	0.878 (0.855 ÷ 0.901)	0.813 (0.588 ÷ 1.038)	Performance $_{est.}$, positive chemicals (sensitivity)	0.878 (0.855 ÷ 0.901)
p-value, sensitivity	< 1.0E-10	9.4E-8	*p-value*, sensitivity	< 1.0E-10
Performance $_{est.}$, negative chemicals (specificity)	0.824 (0.810 ÷ 0.838)	0.806 (0.678 ÷ 0.934)	Performance $_{est.}$, negative chemicals (specificity)	0.824 (0.810 ÷ 0.838)
p-value, specificity	< 1.0E-10	2.6E-5	*p-value*, specificity	< 1.0E-10

[1)] Confidence ranges and *p-value* are calculated at 95% confidence level

Method 2. *Monte Carlo* cross-validation

In ***Monte Carlo cross-validation*** the original training set is split randomly into training and test set. The advantage of this method (compared to *k-fold* cross validation) is that the proportion between training and test sets does not depend on the number of repetitions in the internal validation procedure. The ***Monte Carlo cross-validation*** (similarly to the *bootstrapping*) suppose creating a large number of new training/test sets (1000 – 10000). Results from application of this statistical method are provided in Table 4.

Table 4. Results from Monte Carlo cross-validation (1000 repetitions).

	75% training set		**63% training set**	
	Training sets	**Test sets**		**Training sets**
Unique chemicals, %	75.0 (75.0 ÷ 75.0)	25.0 (25.0 ÷ 25.0)	Unique chemicals, %	75.0 (75.0 ÷ 75.0)
Performance $_{est.}$, all predictions (accuracy)	0.846 (0.836 ÷ 0.857)	0.830 (0.796 ÷ 0.864)	Performance $_{est.}$, all predictions (accuracy)	0.846 (0.836 ÷ 0.857)
p-value, accuracy	< 1.0E-10	< 1.0E-10	*p-value*, accuracy	< 1.0E-10
Performance $_{est.}$, positive chemicals (sensitivity)	0.878 (0.862 ÷ 0.893)	0.836 (0.782 ÷ 0.890)	Performance $_{est.}$, positive chemicals (sensitivity)	0.878 (0.862 ÷ 0.893)
p-value, sensitivity	< 1.0E-10	< 1.0E-10	*p-value*, sensitivity	< 1.0E-10
Performance $_{est.}$, negative chemicals (specificity)	0.824 (0.809 ÷ 0.839)	0.823 (0.778 ÷ 0.868)	Performance $_{est.}$, negative chemicals (specificity)	0.824 (0.809 ÷ 0.839)
p-value, specificity	< 1.0E-10	3.5E-8	*p-value*, specificity	< 1.0E-10

[1]) Confidence ranges and *p-value* are calculated at 95% confidence level.

6.10. Robustness - Statistics obtained by Y-scrambling

Not performed

6.11. Robustness - Statistics obtained by bootstrap

In bootstrapping a newly derived training set is populated from the original training set of the model by random sampling with replacement until the size of the new training set reaches the size of the original training set. The data not selected for the new training set becomes the new test set. On average, about 63% of original training set data goes into the new training set (some data appear more than once) and 37% remains in the new test set. One of the advantages of this

method is that the new training sets and the original training set are equally sized. The process is repeated many times and the average results are provided in Table 5.

Table 5. Results from bootstrapping (1000 repetitions).

	Training sets	**Test sets**
Unique chemicals, %	63.2 (61.6 ÷ 64.9)	36.8 (35.1 ÷ 38.4)
Performance $_{est.}$, all predictions (accuracy)	0.846 (0.827 ÷ 0.866)	0.828 (0.800 ÷ 0.856)
p-value, accuracy	< 1.0E-10	< 1.0E-10
Performance $_{est.}$, positive chemicals (sensitivity)	0.878 (0.851 ÷ 0.905)	0.831 (0.787 ÷ 0.875)
p-value, sensitivity	< 1.0E-10	< 1.0E-10
Performance $_{est.}$, negative chemicals (specificity)	0.824 (0.798 ÷ 0.850)	0.824 (0.790 ÷ 0.858)
p-value, specificity	< 1.0E-10	1.6E-10

[1]) Confidence ranges and *p-value* are calculated at 95% confidence level.

6.12. Robustness - Statistics obtained by other methods

Not performed

6.13. Comment on the internal validation of the model

The difference between performances of training and test sets - which is a measure for optimism in goodness-of-fit, - is around 0.018 for all predictions, 0.05 for positive chemicals (sensitivity) and 0.007 for negative chemicals (specificity). This difference is very low for negative chemicals, but is higher for positive chemicals and the GOF optimism should be considered.

An explanation for the latter is the fact that the training set of “Skin sensitization” model is relatively small (below 1,500 chemicals) and also is **not well balanced** – 60% of training chemicals are negative which favors training of negative chemicals, but is unfavorable for training positive ones.

The model displays good predictive performance for both positive chemicals (sensitivity) – 88% for training set, around 83% expected sensitivity for external sets, - and for negative chemicals (specificity) – 82% for training set, around 82% expected sensitivity for external sets, - and shows **balanced prediction ability** of the model.

Section 7. Defining predictivity (external validation) – OECD Principle 4

7.1. Availability of the external validation set

External validation of the model was done by using a set of 40 chemicals for which new data were generated in the local lymph node assay (LLNA). This set of chemicals was selected based on the following analysis:

1. The European Inventory of Existing Commersial chemical Substances (ELINCS) was screened through TIMES SS to identify the subset of chemicals which fell within the applicability domain of the model.
2. From this filtered inventory, a set of 160 chemicals that covered a similar chemical distribution as the original training set was chosen using the ChemPick software.
3. A further filter was carried out such that only commercially available chemicals remained; 40 chemicals were then picked at "random" to ideally result in a reasonable balance of predicted sensitisers and non sensitisers.

7.2. Available information for the external validation set

The 40 chemicals were all sourced from Sigma-Aldrich and with at least a purity of 97%. The LLNA as described in OECD 429 was performed.

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

Not available

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

Not available

7.5. Other information about the external validation set

The set with 40 chemicals were tested with LLNA. Test concentrations were selected based on the outcomes of specific range finding studies and to an extent on expert opinion. Details for the data used in the current external validation are available in the corresponding publication:

Patlewicz, G., Dimitrov, S., Low, L., Kern, P., Dimitrova, G., Comber, M., Aptula, A., Philips, R., Niemela, J., Madsen, C., Wedebye, E., Roberts, D., Bailey, P., Mekenyan, O. TIMES-SS – A promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. *Regulatory Toxicology and Pharmacology* 48 (2007) 225-239.

7.6. Experimental design of test set

The external validation set contains substances for which the final experimental outcome calls are consensus agreements achieved through discussion between experts taken in conjuction with the study outcome reported by the testing laboratory. In total, there are:

- 24 non sensitisers and
- 16 sensitisers

7.7. Predictivity – Statistics obtained by external validation

Performance for the 35 chemicals remaining after elimination of the five chemicals that had associated uncertainty in experimental data was found to be 85%.

- Sensitivity: 82%
- Specificity: 88%

7.8. Predictivity – Assessment of the external validation set

As part of the external validation, new data were generated for 40 chemicals in the murine local lymph node assay (LLNA). Syngenta Central Toxicology Laboratory (CTL) was identified as the testing laboratory since it met a number of the criteria in terms of excellence.

7.9. Comment on the external validation of the model

The overall concordance of TIMES Skin sensitization model was found to be 85%. The model predicted 88% of non sensitisers correctly and 82% of the sensitisers correctly.

Section 8. Providing a mechanistic interpretation – OECD Principle 5

8.1. Mechanistic basis of the model

The TIMES Skin sensitization model integrates a simulator of skin metabolism together with a list of alerts for protein binding built based on the training set of 1405 chemicals. Mechanistic rational of each alert is provided by experts based on significant reference support from the literature.

8.2. *A priori* or *a posteriori* mechanistic interpretation

The model building followed the traditional approach:

a. Building a hypothesis for the modelled event,
b. Defining the alerting groups based on parent structures,
c. Fitting of model variable to the observed data,
d. Verification of model quality,

e. Depending on the results found in step *d* model building could continue with step *a, b* or *f,*
f. Determination of the applicability domain and practical application of the model.

8.3. Other information about the mechanistic interpretation

Additional information about the mechanistic interpretation could be fund in Section 2 (2.7).

Section 9. Miscellaneous information

9.1. Comments

Model predictions are fully transparent. The user is able to analyse the whole prediction process and to verify whether it concise with his/her knowledge or purposes.

9.2. Bibliography

Additional references are not provided.

9.3. Supporting information

Additional supporting information is not provided.