PATROLS
Advanced Tools for NanoSafety Testing

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 760813

PATROLS Standard Operating Procedures (SOP)

AOP-anchored Nano-QSAR to predict lung tissue inflammation induced by MWCNTs

This is a SOP recommended for external use by PATROLS

Adapted from the NanoImpactNet SOP, Clift *et al* (Deliverable 5.4 under the European Commission's 7th Framework Programme, Grant Agreement 218539).

Authored by:

Karolina Jagiello, QSAR Lab Ltd. (Poland) k.jagiello@qsarlab.com

Sabina Halappanavar, Health Canada (Canada) sabina.halappanavar@canada.ca

Andrew Williams, Health Canada (Canada) Andrew.williams@canada.ca

Ulla Vogel, The National Research Centre for the Working Environment (Denmark) ubv@nfa.dk

Beata Judzińska, QSAR Lab Ltd. (Poland) b.judzinska@qsarlab.com

Anna Rybińska-Fryca, QSAR Lab Ltd. (Poland) a.rybinska@qsarlab.com

Tomasz Puzyn, QSAR Lab Ltd. (Poland) t.puzyn@qsarlab.com

This SOP is based on the publication: Jagiello, K., Halappanavar, S., Rybińska-Fryca, A., Willliams, A., Vogel, U., Puzyn, T. (2021)'Transcriptomics-Based and AOP-Informed Structure–Activity Relationships to Predict Pulmonary Pathology Induced by Multiwalled Carbon Nanotubes'. *Small*, doi: 10.1002/smll.202003465

Reviewed by:

Document History:

Version	Approval Date	Description of the change	Author(s) of change
1.0	15/04/2021	Initial Document	

Table of Contents

1 Introduction:

DOMAIN: Risk assessment

Conventional toxicity tests used to the risk assessment of substances are mainly based on the measurement of the phenotypical response of test cells or the organism after exposure to a given stressor (Saber 2019). The evidence of the toxicity of nanomaterials occurs in the literature but examining each possible variant of such substances is impossible, due to their wide variation. The identification of the structural properties responsible for the mechanisms of toxicity will allow the development of in silico models to predict toxic effects, leading to a reduction of numbers of required experimental tests. Models that predict the toxicity based on the physicochemical properties of nanomaterials are called Nano-QSAR models (Puzyn 2011). The next step in their development is to use early biological changes as the endpoint, that are part of the toxic mechanism observed in the organism.

This document includes a description of the setup and standard operation procedure for developing the AOP-anchored Nano-QSAR model. The model has been used to predict transcriptomic pathway level response for mice lung fibrosis exposed to different multiwalled carbon nanotubes (MWCNTs). The pathway “agranulocyte adhesion and diapedesis”, which is perturbed across the MWCNTs panel, shows dose-response (Benchmark dose, BMDs), and is anchored to the key events (KEs) identified in the lung fibrosis adverse outcome pathway (AOP173); is considered in the modelling. The developed Nano-QSAR model predicts the BMDL level based on the aspect ratio of MWCNTs.

The model details are presented by using the QSAR model reporting format (QMRF).

1.1 Scope and limits of the protocol

The scope of the research protocol includes the use of the considered AOP-anchored Nano-QSAR model to predict BMDL leading to inflammation in mice (perturbation in agranulocyte adhesion and diapedesis pathway) induced by MWCNTs based on their aspect ratio (κ). The inflammation is observed as one of an early event in lung fibrosis caused by nanotubes (AOP173).

BMDL is an equivalent to the NOAEL and is applied to report point of departures (Halappanavar 2019). On the basis of the determined BMDLs, it is possible to determine the potency with which MWCNTs caused perturbations in the analysed pathways.

The described predictive model is dedicated for MWCNTs. For the reliable predictions of the developed model, it should be determined whether the tested MWCNTs belong to its applicability domain. The presented model uses the leverage approach, to compare the similarity of each nanomaterial to the training set. The value of the threshold leverage is 0,43. All formulas and calculations are presented in 5.8.1.

2 Terms and Definitions:

Nanoscale
Length range approximately from 1 nm to 100 nm

Note 1 to entry: Properties that are not extrapolations from larger sizes are predominantly exhibited in this length range.

[SOURCE : ISO/TS 80004-1: 2016, definition 2.1]

Nanotechnology
Application of scientific knowledge to manipulate and control matter predominantly in the *nanoscale* to make use of size- and structure-dependent properties and phenomena distinct from those associated with individual atoms or molecules, or extrapolation from larger sizes of the same material.

Note 1 to entry: Manipulation and control includes material synthesis.

[SOURCE: ISO/TS 80004-1: 2016, definition 2.3]

Nanomaterial
Material with any external dimension in the *nanoscale* or having internal structure or surface structure in the nanoscale.

Note 1 to entry: This generic term is inclusive of *nano-object* and *nanostructured material*.

[SOURCE: ISO/TS 80004-1: 2016, definition 2.4]

Nano-object
Discrete piece of material with one, two or three external dimensions in the *nanoscale*.

Note 1 to entry: The second and third external dimensions are orthogonal to the first dimension and to each other.

[SOURCE: ISO/TS 80004-1: 2016, definition 2.5]

Nanostructure
Composition of inter-related constituent parts in which one or more of those parts is a *nanoscale* region.

Note 1 to entry: A region is defined by a boundary representing a discontinuity in properties.

[SOURCE: ISO/TS 80004-1: 2016, definition 2.6]

Nanostructured material
Material having internal *nanostructure* or surface nanostructure.

Note 1 to entry: This definition does not exclude the possibility for a *nano-object* to have internal structure or surface structure. If external dimension(s) are in the *nanoscale*, the term nano-object is recommended.

[SOURCE: ISO/TS 80004-1: 2016, definition 2.7]

Engineered nanomaterial
Nanomaterial designed for specific purpose or function

[SOURCE: ISO/TS 80004-1: 2016, definition 2.8]

Nano-QSAR

Method for modelling the relationships between properties of nanomaterial and its toxicity response. [2]

Aspect ratio (κ)

Length-to-diameter ratio of the MWCNT.

Training dataset

A dataset used to development of the model.

Validation dataset

A dataset used to calibration of the model.

3 Abbreviations:

MWCNT – multiwalled carbon nanotube

QSAR – quantitative structure-activity relationship

Nano-QSAR – QSAR model for nanomaterials

AD – applicability domain

BMDL – the lowest benchmark dose level

$BMDL_{AA}$ – the lowest benchmark dose level for agranulocyte adhesion and diapedesis pathway

$BMDL_{AA\_pred}$ – predicted value of the lowest benchmark dose level for agranulocyte adhesion and diapedesis pathway

NOAEL- No Observed Adverse Effect Level

KE – key event

AOP – adverse outcome pathway

MLR – Multiple Linear Regression

R^2 – determination coefficient

$RMSE_C$ – root mean square error of calibration

Q^2_{CV} – cross-validated correlation coefficient

$RMSE_{CV}$ – cross-validated root mean square error of prediction

$Q_{2EXT(F2)}$ – the externally validated determination coefficient

$RMSE_{EXT}$ – the root mean square error of prediction

4 Principle of the Method:

The protocol is based on the Nano-QSAR model, which use the properties of MWCNTs to predict organism response at the level of transcriptomic pathways related to lung tissue inflammation – response anchored in AOP173. The AOP-informed Nano-QSAR model shows a linear relationship between BMDLs and the aspect ratio of MWCNTs values. Its development was based on a consensus model that combines the results of several models with the use of different training subsets.

5 Description of the Method:

5.1 Biological setting & test system used:

Mathematical Multiple Linear Regression (MLR) model was developed on the data collected as a result of lab-based experiment. MLR is a simple, transparent technique in which the response y (here: BMDL) is expressed as a linear combination of independent variables x_i (here: descriptors of structural features):

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_ix_i$$

where b_i are regression coefficients and b_0 is an intercept.

5.2 Chemicals and reagents used:

Not applicable.

5.3 Apparatus and equipment used:

In order to perform the Nano-QSAR modelling analysis, appropriate machine learning languages are required, like R, Python, MATLAB or with a ready-to-use software developed for QSAR.

5.4 Reporting of protected elements:

Not applicable.

5.5 Health and safety precautions:

Prior to any use of this SOP a full risk assessment should be completed, considering all potential risks associated with chemicals equipment and use, in compliance with national regulation. Training of personnel should be completed before any person is working with the SOP.

5.6 Applicability:

The SOP is applicable and has been demonstrated for AOP-anchored Nano-QSAR model for predicting agranulocyte adhesion and diapedesis pathway perturbated by MWCNT following mice inhalation.

All MWCNTs used in the process of modelling are presented in the table below.

Table 1 Properties of MWCNTs used in the presented Nano-QSAR model.

MWCNT name	Length [nm]	Diameter [nm]	Surface area [m^2/g]	Composition [%]	OH [mmol/g]	COOH [mmol/g]	$BMDL_{AA}$ [µg/mouse]
NRCWE-006	5700	65	26.00	99.00	0.08	0.04	8.69
NM-401	4050	20	140.00	99.70	0.03	0.02	0.73
NRCWE-048	1604	15.08	185.00	98.80	0.58	0.29	3.85
NRCWE-045	1553	28.07	119.00	96.30	0.63	0.31	13.02
NRCWE-044	1330	32.55	74.00	98.60	0.23	0.11	15.45
NRCWE-026	850	11	245.80	84.40	0.79	0.40	9.80
NRCWE-043	771.3	26.73	82.00	98.50	0.18	0.09	12.12
NRCWE-049	731.1	13.85	199.00	98.80	0.33	0.16	11.24
NRCWE-046	717.2	17.22	223.00	98.70	0.63	0.32	12.43
NRCWE-047	532.5	12.96	216.00	98.70	0.26	0.13	9.53

Length/diameter – measured by SEM
Surface area – measured according to the Brunauer-Emmet-Teller method
Composition – total content of carbon in the MWCNT measured by CEA
OH and COOH – amount of functionality measured by CEA (assuming that all oxygen was OH- or COOH-group)
$BMDL_{AA}$ – BMDL values for agranulocyte adhesion and diapedesis pathway

5.7 Reagent preparation:

Based on the length and diameter of each MWCNT their aspect ratios (κ) were calculated using the following formula:

$$\kappa=\lambda/\phi$$

where λ is the MWCNT's length, whereas ϕ - nanotube's diameter.

5.8 Procedure:

First, from the canonical pathways perturbated by MWCNTs (Halappanavar 2019) that showed dose-response, for which BMDL value was calculated, pathways that are commonly perturbed across all 10 MWCNT types and associated with the MIE or the KEs identified in the AOP 173 for lung fibrosis, were selected for application in Nano-QSAR modelling. The pathway with the highest linear relationship with analysed structural properties (analysis based on the determination coefficient) was selected for modelling.

Then, the data was splited into the training and external validation set. At first, the 10 MWCNTs were sorted based on the increasing endpoint ($BMDL_{AA}$) value. Compounds with the highest and the lowest endpoint values ($BMDL_{AA}$) were arbitrarily assigned to the training set. Then, three compounds were randomly assigned to the validation set. In effect, the points from validation set were evenly distributed within the range of the endpoint ($BMDL_{AA}$) of the training set nanomaterials. The splitting was repeated several times, in order to obtain various training and validation sets (Table 2).

Table 2 Data splits used in development of the model for $BMDL_{AA}$ (T – training set, V – validation set)

MWCNT	Split 0	Split 1	Split 2	Split 3
NM-401	T	T	T	T
NRCWE-048	V	V	T	V
NRCWE-006	T	T	V	T
NRCWE-047	T	T	T	T
NRCWE-026	V	V	T	T
NRCWE-049	T	T	V	V
NRCWE-043	T	T	T	T
NRCWE-046	T	V	T	T
NRCWE-045	V	T	V	V
NRCWE-044	T	T	T	T

All splits were used to develop the consensus regression model and verify its predictive ability. For each version of the training set, multiple linear regression (MLR) was applied as the method of supervised modelling.

Set of models developed for each splitting pattern:

Model 0 (Split 0):

$$BMDL_{AA} = 15.01 - 0.071 \times \kappa$$

Model 1 (Split 1):

$$BMDL_{AA} = 15.31 - 0.072 \times \kappa$$

Model 2 (Split 2):

$$BMDL_{AA} = 14.92 - 0.075 \times \kappa$$

Model 3 (Split 3):

$$BMDL_{AA} = 15.06 - 0.071 \times \kappa$$

Using the results of that several models, the consensus model was created. The split into test and validation set, and model equation are presented below.

Table 3 Data split used in the consensus model for $BMDL_{AA}$ (T – training set, V – validation set)

MWCNT	Split
NM-401	T
NRCWE-006	T
NRCWE-047	T
NRCWE-049	T
NRCWE-043	T
NRCWE-046	T
NRCWE-044	T
NRCWE-048	V
NRCWE-026	V
NRCWE-045	V

Consensus model equation:

$$BMDL_{AA} = 15.07 - 0.07 \times \kappa$$

5.8.1 Testing for nanomaterial interference:

In order to assess the plausibility of the predictions, it is necessary to examine whether the investigated nanoparticles belong to the applicability domain (AD) of the model. One of the methods to establish border of AD is presented here the leverage approach. To compare the similarity of each nanomaterial to the training set, is used value of the leverage, calculated according to the following equation:

$$h_i = x_i^T (\mathbf{X}^T \mathbf{X})^{-1} x_i$$

where x_i is the vector of descriptors calculated for the considered nanomaterial and **X** is the matrix of descriptors calculated for all MWCNTs from the training set.

The obtained value is compared with the threshold leverage value (h *), which is the boundary of the AD. It is calculated as h* = 3p'/n, where p' is the number of descriptors in equation plus one, and n is the number of nanomaterials in the training set. Furthermore, MWCNTs with residuals differing by more than +/- 3 standard deviations are not equated. In the table 4 are presented standardized residuals and the leverage values for every nanomaterial. The graphical representation of AD has been visualized in the Figure 1.

Table 4 Standardized residuals and the leverage values for the model predicting $BMDL_{AA}$.

MWCNT	Residuals	Leverages
NM-401	-0.01	0.92
NRCWE-006	0.00	0.16

NRCWE-047	-1.34	0.18
NRCWE-049	0.05	0.16
NRCWE-043	-0.42	0.22
NRCWE-046	0.27	0.18
NRCWE-044	1.90	0.18
NRCWE-048	-1.57	0.20
NRCWE-026	0.18	0.15
NRCWE-045	0.93	0.15

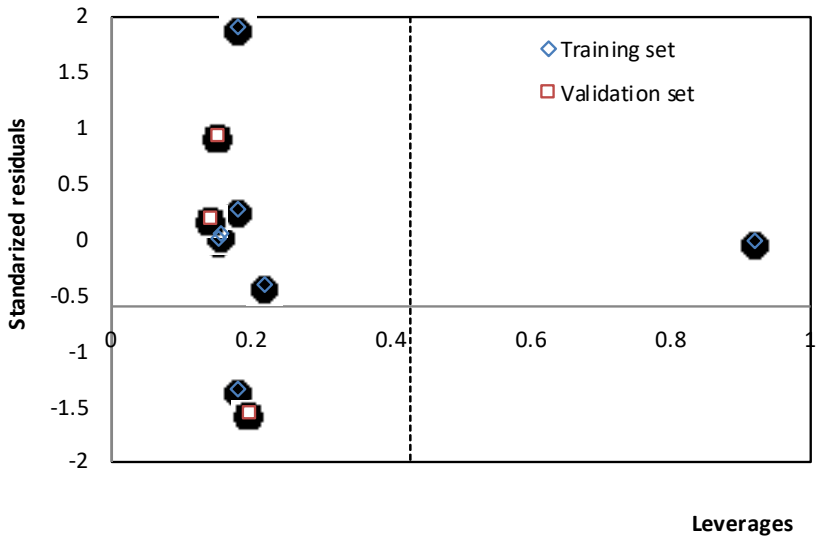

Figure 1 Applicability domain (leverage approach) for model allowing to predict BMDL (dashed line expresses the h value)*

5.9 Quality control & acceptance criteria:

Based on the $BMDL_{AA}$ values predicted by the model and the values obtained experimentally, the predicted vs. observed plot were created for model quality evaluation. The result of the analysis for the presented models is presented below.

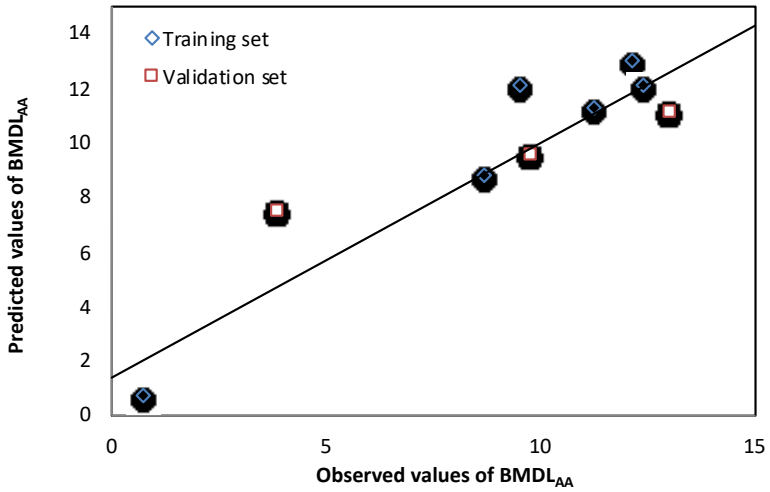

Figure 2 Predicted vs. observed values of $BMDL_{AA}$ of investigated MWCNTs.

The straight line represents perfect agreement between experimental and calculated values. Blue diamonds represent the training set values and squares with red frame concern values from the validation set. The distance of each symbol from the straight line corresponds to its deviation from the related experimental value.

In the Table 5, the values of the used descriptor, experimentally obtained and predicted BMDLs are summarized.

Table 5 Details of the descriptor, observed and predicted BMDL values

MWCNT	κ	$BMDL_{AA}$	$BMDL_{AA\_pred}$
NM-401	202.50	0.73	0.75
NRCWE-006	87.69	8.69	8.84
NRCWE-047	41.09	9.53	12.12
NRCWE-049	52.79	11.24	11.30
NRCWE-043	28.86	12.12	12.98
NRCWE-046	41.65	12.43	12.08
NRCWE-044	40.86	15.45	12.14
NRCWE-048	106.37	3.85	7.52
NRCWE-026	77.27	9.80	9.57

NRCWE-045	55.33	13.02	11.12

6 Data Analysis and Reporting of Data:

The goodness-of-fit of the models were assessed using determination coefficient (R^2) and Root Mean Square Error of Calibration ($RMSE_C$):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{pred}\right)^2}{\sum_{i=1}^{n}\left(y_i^{obs} - \tilde{y}_i^{obs}\right)^2}$$

$$RMSE_C = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{pred}\right)^2}{n}}$$

where: y_i^{obs} – the experimental (observed) value of the property for the *i*th MWCNT; y_i^{pred} – the predicted value for the *i*th MWCNT; n – the number of nanoparticles in the training set.

The robustness of the model on the presence or absence of particular MWCNT in the training set the cross-validation leave-one-out method (CV_{LOO}) were applied. For this purpose, the values of the cross-validated correlation coefficient Q^2_{CV} and cross-validated root mean square error of prediction $RMSE_{CV}$ were calculated. Moreover, the F-test were tested to indicate the presence of influential points. All equations are presented below:

$$Q^2_{CV} = 1 - \frac{\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{pred_{CV}}\right)^2}{\sum_{i=1}^{n}\left(y_i^{obs} - \tilde{y}_i^{obs}\right)^2}$$

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{pred_{CV}}\right)^2}{n}}$$

$$F = \frac{1 - Q^2_{CV}}{1 - R^2}$$

where: y_i^{obs} – the experimental (observed) value of the property for the *i*th MWCNT; y_i^{predCV} – predicted value for the temporary excluded (cross-validated) *i*th MWCNT; $\tilde{y}^{obs}$ – the mean experimental value of the modeled property in the training set; n – the number of nanoparticles in the training set.

Other parameters that were calculated in order to assess the quality of the model are the externally validated determination coefficient $Q^2_{EXT(F2)}$ and the root mean square error of prediction $RMSE_{EXT}$. They enable to assess the ability of the model to predict the endpoint value of MWCNTs from outside training set.

$$Q^2_{EXT(F2)} = 1 - \frac{\sum_{j=1}^{k}\left(y_j^{obs} - y_j^{pred}\right)^2}{\sum_{j=1}^{k}\left(y_j^{obs} - \tilde{y}_j^{obs}\right)^2}$$

$$RMSE_{EXT} = \sqrt{\frac{\sum_{j=1}^{k}\left(y_j^{obs} - y_j^{pred}\right)^2}{k}}$$

where: y_j^{obs} – the experimental value of the property for the j^{th} MWCNT from validation set; y_j^{pred} – the predicted value for j^{th} MWCNT; $\tilde{y}_j^{obs}$ – the mean experimental value of the property in the validation set; k – the number of nanoparticles in the validation set.

The statistics of the models are presented in the table below:

R^2	$RMSE_C$	Q^2_{CV}	$RMSE_{CV}$	$Q^2_{EXT(F2)}$	$RMSE_{EXT}$	F
Model 0						
0.85	1.63	0.78	1.99	0.69	2.39	1.47
Model 1						
0.84	1.76	0.74	2.22	0.74	2.23	1.63
Model 2						
0.81	2.07	0.47	3.47	0.92	1.34	2.79
Model 3						
0.85	1.63	0.77	2.03	0.68	2.4	1.53
Consensus model						
0.86	1.63	-	-	0.62	2.34	-

7 Publications:

Jagiello, K., Halappanavar, S., Rybińska-Fryca, A., Willliams, A., Vogel, U., Puzyn, T. (2021)'Transcriptomics-Based and AOP-Informed Structure–Activity Relationships to Predict Pulmonary Pathology Induced by Multiwalled Carbon Nanotubes'. *Small*, doi: 10.1002/smll.202003465

8 References

Halappanavar, S., Rahman, L., Nikota, J., Poulsen, S. S., Ding, Y., Jackson, P., Wallin, H., Schmid, O., Vogel, U., Williams, A. (2019) 'Ranking of nanomaterial potency to induce pathway perturbations associated with lung responses', *NanoImpact*, 14, pp. 100158-100175.

Saber, A.T., Poulsen, S.S., Hadrup, N., Jacobsen, N.R., Vogel, U.(2019) 'Commentary: the chronic inhalation study in rats for assessing lung cancer risk may be better than its reputation', *Particle and Fibre Toxicology*, 16, pp.16-44

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T.P., Michalkova, A., Hwang, H.-M., Toropov, A., Leszczynska, D., Leszczynski, J. (2011) 'Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles', *Nature Nanotechnology*, 6, pp. 175-178.

9 APPENDIX – QSAR Model Reporting Format

The QSAR Model Reporting Format (QMRF) of the AOP-anchored Nano-QSAR model predicting transcriptomic pathway level response for lung tissue inflammation of a set of multiwalled carbon nanotubes (MWCNTs).

1. QSAR identifier
1.1. QSAR identifier (title):
Transcriptomics-Based and AOP-Informed Structure–Activity Relationships to Predict Pulmonary Pathology Induced by Multiwalled Carbon Nanotubes
1.2. Other related models:
-
1.3. Software coding the model:
-
2. General information
2.1. Date of QMRF:
October, 2020
2.2. QMRF author(s) and contact details:
E. Wyrzykowska, QSAR Lab Ltd, e.wyrzykowska@qsarlab.com Dr K. Jagiełło, QSAR Lab Ltd, k.jagiello@qsarlab.com
2.3. Date of QMRF update(s):
-
2.4. QMRF update(s):
-
2.5. Model developer(s) and contact details:
Dr K. Jagiełło, QSAR Lab Ltd, k.jagiello@qsarlab.com Dr S. Halappanavar, Health Canada (Canada) sabina.halappanavar@canada.ca A. Rybińska-Fryca, QSAR Lab Ltd, a.rybinska@qsarlab.com A. Williams, Health Canada (Canada) Andrew.williams@canada.ca Prof. U. Vogel, The National Research Centre for the Working Environment (Denmark) ubv@nfa.dk Prof. T.Puzyn, QSAR Lab Ltd, t.puzyn@qsarlab.com
2.6. Date of model development and/or publication:
2021
2.7. Reference(s) to main scientific papers and/or software package:
Jagiello, K. *et al.* Transcriptomics-based and AOP-informed structure-activity relationships to predict pulmonary pathology induced by multiwalled carbon nanotubes. *Small* (2021) doi: *10.1002/smll.202003465*

2.8. Availability of information about the model:

The model is non-proprietary. The algorithm and datasets are provided. Detailed information available in the original paper and Supporting Information:

Jagiello, K. *et al.* Transcriptomics-based and AOP-informed structure-activity relationships to predict pulmonary pathology induced by multiwalled carbon nanotubes. *Small* (2020) doi: *10.1002/smll.202003465*

2.9. Availability of another QMRF for exactly the same model:

None to date.

3. Defining the endpoint – OECD Principle 1

3.1. Species:

C57BL/6 mice

3.2. Endpoint:

Other (QMRF 6.6.)

3.3. Comment on the endpoint:

The Nano-QSAR model predicts transcriptomic pathway level response for mice lung fibrosis exposed to different multiwalled carbon nanotubes (MWCNTs). The pathways “agranulocyte adhesion and diapedesis”, that is perturbed across the MWCNTs panel, shows dose response (Benchmark dose, BMDs), and is anchored to the key events (KEs) identified in the lung fibrosis adverse outcome pathway (AOP); is considered in the modelling. The developed Nano-QSAR model predicts the BMDs level (BMDL) based on the aspect ratio of MWCNTs.

Gene expression profiles from lung tissues of female adult C57BL/6 mice (N = 5 per group) intratracheally exposed at dose levels 18, 54, and 162 µg per mouse, or 6, 18, and 54 µg per mouse of individual MWCNTs and control mice exposed to vehicle only, were analysed to identify the differentially expressed genes and pathways perturbed following exposure to MWCNTs.

3.4. Endpoint units:

µg per mouse

3.5. Dependent variable:

The dependent variable is the experimentally measured, non-transformed, BMDL response.

3.6. Experimental protocol:

The detailed methodology involving animal exposure and sample collection, generation of lung transcriptomics data, and bioinformatics analysis of the expression changes in genes and pathways was previously published:

[1] S. S. Poulsen, A. T. Saber, A. Williams, O. Andersen, C. Kobler, R. Atluri, M. E. Pozzebon, S. P. Mucelli, M. Simion, D. Rickerby, A. Mortensen, P. Jackson, Z. O. Kyjovska, K. Molhave, N. R. Jacobsen, K. A. Jensen, C. L. Yauk, H. Wallin, S. Halappanavar, U. Vogel, Toxicol. Appl. Pharmacol. 2015, 284, 16.

[2] S. S. Poulsen, N. R. Jacobsen, S. Labib, D. Wu, M. Husain, A. Williams, J. P. Bogelund, O. Andersen, C. Kobler, K. Molhave, Z. O. Kyjovska, A. T. Saber, H. Wallin, C. L. Yauk, U. Vogel, S. Halappanavar, PLoS One 2013, 8, e80452.

[3] S. S. Poulsen, P. Jackson, K. Kling, K. B. Knudesn, V. Skaug, Z. O. Kyjovska, B. L. Thomsen, P. A. Clausen, R. Atluri, T. Berthing, S. Bengston, H. Wolff, K. A. Jensen, H. Wallin, S. Halappanavar, U. Vogel, Nanotoxicology 2016, 10, 1263.

3.7. Endpoint data quality and variability:

The used experimental data (training and validation set) were obtained in the same laboratory. The data (BMDL responses) were previously published:

S. Halappanavar, L. Rahman, J. Nikota, S. S. Poulsen, Y. Ding, P. Jackson, H. Wallin, O. Schmid, U. Vogel, A. Williams, NanoImpact 2019, 14, 100158.

4. Defining the algorithm – OECD Principle 2

4.1. Type of model:

Nano-QSAR MLR consensus model

4.2. Explicit algorithm:

Nano-QSAR consensus model

Multiple linear regression (MLR) model

The developed model is based on the MWCNTs' aspect ratio (descriptor).

The details are available in original paper and Supporting Information:

Jagiello, K. *et al.* Transcriptomics-based and AOP-informed structure-activity relationships to predict pulmonary pathology induced by multiwalled carbon nanotubes. *Small* (2020) doi: *10.1002/smll.202003465*

4.3. Descriptors in the model:

The Nano-QSAR model is based on the one descriptor, MWCNTs' aspect ratio (length to diameter ratio)

4.4. Descriptor selection:

By expert knowledge, supported with correlation value with modelled parameter.

The initial number of considered MWCNTs' descriptors – 7.

The MWCNTs descriptors reflected length, diameter, surface area, aspect ratio, composition and surface modifications with OH or COOH functional groups.

4.5. Algorithm and descriptor generation:

Descriptors are derived from experimental characterization.

4.6. Software name and version for descriptor generation:

-

4.7. Chemicals/ Descriptors ratio:

7 (training) and 3 (validation) MWCNTs / 1 descriptor

5. Defining the applicability domain – OECD Principle 3

5.1. Description of the applicability domain of the model:

The applicability domain is defined with the leverage approach. The leverage values (h_i) reflect the similarity of particular compounds to the training set based on their values of aspect ratio descriptor. Border of the applicability domain is determined by the threshold leverage value (h*), which is calculated as h* = 3p'/n, where p' is the number of descriptors in equation plus

one, and n is the number of compounds in the training set; and residuals thresholds differing by more than +/- 3 standard deviations.

5.2. Method used to assess the applicability domain:

Leverage approach.

5.3. Software name and version for applicability domain assessment:

-

5.4. Limits of applicability:

Multiwalled carbon nanotubes (MWCNTs) structures

Descriptor value range: aspect ratio <28.86; 202.50>

Residuals thresholds differing by more than +/- 3 standard deviations.

6. Defining goodness-of-fit and robustness – OECD Principle 4

6.1. Availability of the training set:

The training set is available in the original paper:

Jagiello, K. *et al.* Transcriptomics-based and AOP-informed structure-activity relationships to predict pulmonary pathology induced by multiwalled carbon nanotubes. *Small* (2020) doi: *10.1002/smll.202003465*

6.2. Availability information for the training set:

a. Chemical names – yes
b. CAS numbers – not applicable
c. SMILES – not applicable
d. InChI codes – not applicable
e. MOL files – no
f. Structural formula – no
g. Any other structural information – yes, available experimental characterization

6.3. Data for each descriptor variable for the training set:

Available

6.4. Data for the dependent variable (response) for the training set:

Available

6.5. Other information about the training set:

Available experimental characterization of the training set.

6.6. Pre-processing of data before modelling:

-

6.7. Statistics for goodness-of-fit:

$R^2 = 0.86$

$RMSE_C = 1.63$

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

-

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

-

6.10. Robustness – Statistics obtained by Y-scrambling:

-

6.11. Robustness – Statistics obtained by bootstrap:

-

6.12. Robustness – Statistics obtained by other methods:

-

7. Defining predictivity – OECD Principle 4

7.1. Availability of the external validation set:

Available

7.2. Availability information for the external validation set:

a. Chemical names – yes
b. CAS numbers – not applicable
c. SMILES – not applicable
d. InChI codes – not applicable
e. MOL files – no
f. Structural formula – no
g. Any other structural information – yes, available experimental characterization

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

Available

7.4. Data for the dependent variable (response) for the external validation set:

Available

7.5. Other information about the external validation set:

Available experimental characterization of the external validation set.

7.6. Experimental design of test set:

-

7.7. Predictivity – Statistics obtained by external validation:

Q^2_{EXT} = 0.62

$RMSE_{EXT}$ = 2.34

7.8. Predictivity – Assessment of the external validation set:

The external validation set is relatively small but representative for the investigated dataset of MWCNTs. The descriptor values characterizing the nanoforms in the external validation set are within the range of descriptor values of nanoforms used in the training set. The distribution of the response values of the nanoforms in the training set and external validation set is comparable.

7.9. Comments on the external validation of the model:

-

8. Providing a mechanistic interpretation – OECD Principle 5

8.1. Mechanistic basis of the model:

The Nano-QSAR model is based on the correlation between the aspect ratio of MWCNTs and the dose response (Benchmark dose level, BMDL) values of the agranulocyte adhesion and

diapedesis pathway. The modelling results indicate that an increasing aspect ratio causes a decrease in the BMDL values. The activation of the pathway is observed at the lowest doses of MWCNTs with the highest aspect ratio values.

8.2. A priori or a posteriori mechanistic interpretation:

The mechanistic interpretation was determined a priori based on the mechanism of action widely discussed in the literature, supported with the analysis of descriptors correlation with the modelled BMDL response.

8.3. Other information about the mechanistic interpretation:

-

9. Miscellaneous information

9.1. Comments:

The developed Nano-QSAR model provides new insights into predicting the AO by quantitatively expressing the changes in structural properties of MWCNTs that are detrimental to inducing perturbation in pathways associated with early upstream KEs essential for the occurrence of the AO.

9.2. Bibliography:

<link to the app>

9.3. Supporting information:

-

10. Summary for the JRC QSAR Model Database (compiled by JRC)

10.1. QMRF number:

10.2. Publication date:

10.3. Keywords:

10.4. Comments: