	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **QSPR model for the heat of decomposition of organic peroxides using constitutional descriptors**
	Printing Date: **25 mars 2020**

1.QSAR identifier

1.1.QSAR identifier (title):

QSPR model for the heat of decomposition of organic peroxides using constitutional descriptors

1.2.Other related models:

1.3.Software coding the model:

2.General information

2.1.Date of QMRF:

25/03/2020

2.2.QMRF author(s) and contact details:

[1]Guillaume FAYET INERIS guillaume.fayet@ineris.fr

[2]Patricia ROTUREAU INERIS patricia.rotureau@ineris.fr

2.3.Date of QMRF update(s):

09/10/2017 - Version 1

25/03/2020 - Version 2

2.4.QMRF update(s):

25/03/2020 - Guillaume FAYET - corrections of the software name in sections 4.5 and 4.6.

2.5.Model developer(s) and contact details:

[1]Guillaume FAYET INERIS guillaume.fayet@ineris.fr

[2]Patricia ROTUREAU INERIS patricia.rotureau@ineris.fr

2.6.Date of model development and/or publication:

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

V. Prana, P. Rotureau, D. Andre, G. Fayet, and C. Adamo, Development of Simple QSPR Models for the Prediction of the Heat of Decomposition of Organic Peroxides, Mol. Inf. 2017, 36, 1700024. http://onlinelibrary.wiley.com/doi/10.1002/minf.201700024/full

2.8.Availability of information about the model:

The model is non proprietary. Training and validation sets are available

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

3.Defining the endpoint - OECD Principle 1

3.1.Species:

3.2.Endpoint:

QMRF 6. Other Heat of decomposition

3.3.Comment on endpoint:

This property is used in the screening procedure for substances which may have explosive properties to classify substances according to the Transport of Dangerous Goods (international regulation) and the Classification, Labelling and Packaging of substances and mixtures (European CLP regulation). Therefore, it can be used in REACH context.

3.4.Endpoint units:

J/g

3.5.Dependent variable:

deltaH/C - heat of decomposition divided par concentration (in %wt)

3.6.Experimental protocol:

The experimental data were collected from differential scanning calorimetry (DSC) (see ref. 2). The calorimetric tests were performed using DSC131 from SETARAM and with DSC821e from Mettler-Toledo, with a scanning rate of 5 K/min from ambient temperature to 300°C. A few milligrams of sample were introduced, at ambient temperature, into a closed stainless steel crucible. The DSC vessels were previously washed, passivated, rinsed and dried.

3.7.Endpoint data quality and variability:

The calorimetric tests were carried out on two different calorimeters with cross-checking, indicating that the results obtained were not equipment-dependent and were within the precision limits of the DSC method recommended in ASTM E537-2.

Each sample was tested three times to establish good reproducibility.

In order to consider concentration effects, three organic peroxide samples (one dialkyl peroxide and two peroxyesters) were diluted with an inert solvent to obtain different concentrations from 7% up to 99% wt.

Linearity of the concentration effect was checked (see ref. 2).

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR - multilinear regression

4.2.Explicit algorithm:

deltaH/C = - 663 nOO - 699 nOOH – 4.79 OB + 11 nsingle - 2036

4.3.Descriptors in the model:

[1]nOO Number of O-O bonds

[2]nOOH Number of O-OH bonds

[3]OB Oxygen Balance according to TDG definition: OB = -1600(2 nC + 0.5 nH - nO)/Mw

[4]nsingle Number of single bonds

4.4.Descriptor selection:

The descriptor selection was performed using the Best Multi Linear Regression (BMLR) approach as implemented in Codessa program. This stepwise approach of descriptor selection starts by reducing the number of descriptors by eliminating those which present an insignificant variance or correlation with the studied property. Then, when two descriptors are highly correlated together, the one presenting the lower correlation with the property is also eliminated. Therefore, two-parameter regressions involving orthogonal descriptors are computed. Higher rank models are then built by adding new descriptors presenting no correlation with the ones already present in the model until no improvement of the model is found. Finally, the algorithm proposes, at each rank (i. e. for each number of descriptors), the model presenting

the highest correlation with the studied property. The final model was chosen as the best compromise between correlation observed within the training set and number of descriptors, taking also into consideration the relevance of the descriptors into the model from a chemical point of view. To further ensure the relevance of the descriptors in the models, t-test values were checked at a 95% confidence level.

4.5.Algorithm and descriptor generation:

Multiple linear regression was computed by Codessa software to generate the model. Descriptors were calculated from examination of chemical structures.

4.6.Software name and version for descriptor generation:

Codessa 2.7.16

Selection of descriptors and multilinear regression calculation

http://www.semichem.com/codessa/default.php

4.7.Chemicals/Descriptors ratio:

37 chemicals / 4 descriptors = 9.25

The ratio for the training set is 25/4 = 6.25

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain of the model is defined in terms of:

- chemical family: organic peroxides, i.e. organic compounds presenting a O-O bond ;
- chemical structure space in the training set based on the descriptors of the model using the method described in section 5.2 ;
- range of property values in the training set.

Only one molecule of the validation set is out of the applicability domain : tert-butyl hydroperoxide.

5.2.Method used to assess the applicability domain:

The applicability domain of the model was built, for each descriptor and for the property, by the range of values represented among the molecules of the training set. The AD ranges for each descriptor and for the property are available in section 5.4.

5.3.Software name and version for applicability domain assessment:

5.4.Limits of applicability:

Ranges of values defining the applicability domain:

- nOO: [1 ; 2]
- nOOH: [0 ; 1]
- OB: [-268.9 ; -131.9]
- ns: [15 ; 103]
- deltaH/C: [-2370 ; -484]

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: Yes

Formula: Yes

INChI: Yes

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:

The training set contains 2/3 of the data, i.e. 25 molecules. To ensure that the validation set covers at best the chemical diversity of the domain of applicability of the model, i. e. of the molecules in the training set, the partitioning was carried out using the score plots issued from Principal Components Analysis (PCA) computed with the R program. Within this analysis, an initial set of 128 constitutional and constitutional descriptors was projected into two principal components (PC), accounting for the maximum variance of structures in the data set. The score plot based on these first two PCs accounts for the maximum variance of the data set in terms of chemical diversity. So, the training and validation molecules were selected to be homogeneously distributed into the global chemical space represented on this score plot and to favour at best the presence of molecules of each family of organic peroxides in both sets. Finally, it has been checked that the distribution of data was also homogeneous in terms of property.

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

n=25; R^2=0.89; MAE=118 J/g; RMSE=166 J/g

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

Q^2LOO=0.79

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

$Q^2$10CV=$Q^2$7CV=0.81; $Q^2$5CV=0.80

6.10.Robustness - Statistics obtained by Y-scrambling:

500 iterations; Average R^2(YS)=0.17 ; Standard Deviation of R^2(YS)=0.10

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

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: Yes

Formula: Yes

INChI: Yes

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:

The training set contains 1/3 of the data, i.e. 12 molecules.

7.6.Experimental design of test set:

Selection of molecules in both sets was performed as described in section 6.6.

7.7.Predictivity - Statistics obtained by external validation:

n=12; R^2=0.79; MAE=171 J/g; RMSE=306 J/g; Q^2F1=0.78; Q^2F2=0.78; Q^2F3=0.94; CCC=0.89

in AD: n=11; R^2=0.80; MAE=162 J/g; RMSE=311 J/g; Q^2F1=0.77; Q^2F2=0.77; Q^2F3=0.95; CCC=0.86

7.8.Predictivity - Assessment of the external validation set:

No bias of representativity was identified in the property distribution or in terms of chemical diversity (assessed by a principal component analysis and an inspection of chemical structures).

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 model is focused on organic peroxides that are known to decompose following a same decomposition mechanism involving the breaking of the O-O bond [refs. 3-4].

In the equation, the occurence of nOO and nOOH descriptors is in accordance with this knowledge. Moreover, the presence of the Oxygen Balance is also relevant since it is a pre-selection criterion in the regulatory classification of explosives including organic peroxides.

8.2.A priori or a posteriori mechanistic interpretation:

All descriptors involved in the model were a priori chosen to be included into the set of investigated molecular descriptors since they were identified as possibly relevant for the thermal stability of organic peroxides according to the author's expert knowledge on this family of compounds and on this class of properties.

Their relevance was finaly verified by the statistical analysis since they are found in the final model.

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

[1]V. Prana, P. Rotureau, D. Andre, G. Fayet, C. Adamo, Developement of Simple QSPR Models for the Prediction of the Heat of Decomposition of Organic Peroxides, Molecular Informatics, 2017, 36, 1700024 http://onlinelibrary.wiley.com/doi/10.1002/minf.201700024

[2]V. Prana, P. Rotureau, G. Fayet, D. Andre, S. Hub, P. Vicot, L. Rao, C. Adamo, Prediction of the thermal decomposition of organic peroxides by validated QSPR models, http://www.sciencedirect.com/science/article/pii/S0304389414003446

[3]R. Benassi, U. Folli, S. Sbardellati, F. Taddei, Conformational properties and homolytic bond cleavage of organic peroxides. I. An empirical approach based upon molecular mechanics and ab initio calculations, Journal of Computational Chemistry, 1993, 14, 379-391 http://onlinelibrary.wiley.com/doi/10.1002/jcc.540140402

[4]R. Benassi, F. Taddei, Homolytic bond-dissociation in peroxides, peroxyacids, peroxyesters and related radicals: ab-initio MO calculations, Tetrahedron, 1994, 50, 4795-4810 http://www.sciencedirect.com/science/article/pii/S0040402001850171

9.3.Supporting information:

Training set(s)

train.zip	file:///C:\Users\Fayet\Documents\WORK\Modèles QSPR\Proposition Modèles Toolbox\QMRF POs simple\MOL\train.zip
train-data.pdf	file:///C:\Users\Fayet\Documents\WORK\Modèles QSPR\Proposition Modèles Toolbox\QMRF POs simple\train-data.pdf

Test set(s)

valid.zip	file:///C:\Users\Fayet\Documents\WORK\Modèles QSPR\Proposition Modèles Toolbox\QMRF POs simple\MOL\valid.zip
valid-data.pdf	file:///C:\Users\Fayet\Documents\WORK\Modèles QSPR\Proposition Modèles Toolbox\QMRF POs simple\valid-data.pdf

Supporting information

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

To be entered by JRC

10.4.Comments:

To be entered by JRC