First QSPR models to predict the thermal stability of potential self-reactive substances

Guillaume Fayet, Annett Knorr, Patricia Rotureau

First QSPR models to predict the thermal stability of potential self-reactive substances

Guillaume Fayet[a,*], Annett Knorr[b], Patricia Rotureau[a]

[a]Ineris, Accidental Risk Division, Parc Technologique Alata, 60550 Verneuil-en-Halatte, France

[b]BAM, Bundesanstalt für Materialforschung und -prüfung, Unter den Eichen 87, 12205 Berlin, Germany

*Corresponding author: guillaume.fayet@ineris.fr; tel: +33(0)344618126

Abstract

Self-reactive substances are unstable chemical substances which can easily decompose and may lead to explosion in transport, storage, or process situations. For this reason, their thermal stability properties are required to assess possible process safety issues and for classification purpose. In this study, the first quantitative structure–property relationships (QSPR) dedicated to this class of compounds were developed to predict the heat of decomposition of possible self-reactive substances from their molecular structures. The database used to develop and validate the models was issued from a dedicated experimental campaign on 50 samples using differential scanning calorimetry in homogeneous experimental conditions. QSPR models were derived using the GA-MLR methods (using a genetic algorithm and multi-linear regressions) using molecular descriptors calculated by Dragon software based on two types of inputs: 3D molecular structures determined using the density functional theory (DFT), allowing access to three-dimensional descriptors, and from SMILES codes, favoring the access to simpler models, requiring no preliminary quantum chemical calculations. All models respected the OECD validation guidelines for regulatory acceptability of QSPR models. They were tested by internal and external validation tests and their applicability domains were defined and analyzed.

Keywords: Quantitative Structure-Property Relationships, self-reactive substances, thermal stability, genetic algorithm.

1. Introduction

Self-reactive substances are thermally unstable substances and mixtures liable to undergo self-accelerating decomposition without contribution of oxygen (from air) and even, in some cases, to be explosive under the effect of various types of stimuli (e.g. impact or friction). So, they can cause thermal runaways in chemical process. They can also react dangerously with other substances. Finally, their highly exothermic decomposition can cause fire or explosion during processing, storage, and transport [1, 2].

To favor their safe handling, the self-reactive substances are classified specifically in chemical safety regulations like the UN Recommendations for the Transport of Dangerous Good, Model Regulation [3] (in class 4.1), or the UN Globally Harmonized System of Classification and Labelling of Chemicals (GHS) [4] and the European Regulation related to the Classification, Labelling and Packaging of substances (CLP) [5] (as "Self-reactive substances and mixtures"). The classification is based on a series of experimental tests to evaluate their detonation, deflagration and thermal explosion properties, their response to heating under confinement and their explosive power. In addition, their self-accelerating decomposition (SADT) is assessed to determine the lowest temperature at which a self-accelerating decomposition could occur in the wanted packaging (package size can vary from 50 kg to intermediate bulk containers (IBC) or up to non-isolated or isolated tanks with higher volumes) during storage or transport to decide if a control temperature is necessary (and to define it, if appropriate) [6, 7]. In compliance with the Manual of Tests and Criteria [7] a screening procedure for substances which may be self-reactive substances has to be applied. Criteria are the presence of specified chemical groups, the estimated SADT for a 50 kg package and the decomposition energy, estimated by a suitable calorimetric technique like DSC. If the SADT is higher than 75°C and the decomposition energy is lower than 300 J/g, the classification procedures need not be applied.

It should be noted, that also organic peroxides are thermally unstable substances which can undergo a self-accelerating decomposition, but they are summarized in specific hazard class in chemical safety regulations (identified by the presence of a bivalent -O-O- structure) and are not subject to the present investigation. Self-reactive substances are also characterized by specific chemical groups, whereby its nature has a higher variety compared to organic peroxides.

Here, we focus on the heat of decomposition which represents the heat released by the substance during decomposition and represents a relevant information to identify possible process safety issues due to thermal hazards. It does not only play a role in the screening procedures for potential self-reactive substances, but also for substances, which may have explosive properties and for which, according to the Manual of Tests and Criteria [7], the so-called acceptance procedure for explosives need or not need to be applied. The acceptance procedure is used to determine whether a substance or article as offered for classification is a candidate for the class of explosives (called Class 1 in transport regulation) or not. If the decomposition energy is equal or higher than 500 J/g and the onset temperature of the decomposition is lower than 500 °C, then the acceptance procedure for an organic substance or a homogenous mixture of organic substances need to be applied.

Information on the physical hazards of self-reactive or other hazardous substances are up to now mainly obtained from experimental tests. But these tests are complex, costly and require large quantity of substances and dedicated facilities (notably to access SADT data on real-size packages). In this context, predictive methods or alternative small-scale tests represent a real improvement as an

alternative to huge experimental characterizations. For instance, Fisher and Goetz [8] developed a method to determine the SADT of self-reactive substances based on Accelerating Rate Calorimeter (ARC) data. Malow and Wehrstedt [9] also proposed to calculate the SADT of pure liquid organic peroxides from dynamic DSC runs using a transfer equation based on the standard theory of explosion by Semenov as previously done from DTA runs by Steinbach and Klais [10].

To access physical hazard information for process engineering and process safety analyzes, without any experimental data, Quantitative Structure-Property Relationships (QSPR) offer the possibility to estimate properties based only on the knowledge of their molecular structures. Such models were explicitly mentioned in the REACH regulation [11] among the alternative methods recommended to reduce the cost and time for the registration, evaluation and authorization of chemicals in Europe. To support regulatory applications of QSPR models, validation principles were proposed by the Organization for Economic Co-operation and Development (OECD) [12]. They concern the importance of transparency not only in the definition of the algorithm but also in the definition of endpoints (including experimental protocols) and of the domains of applicability. They also promote the use of appropriate estimations of the performances of models (including an external validation to estimate predictivity). Finally, the mechanistic interpretation of the model is encouraged. Beyond regulatory context, QSPR models allow property estimations for screening purposes in early R&D phases of new substances and the design of new industrial process, even when chemicals are not yet synthesized or when experiments are too expensive or dangerous (e.g. in the case of toxic substances).

QSPR models have been successfully developed for different kinds of properties, from toxicology [13] to physical chemistry [14]. Recent reviews [15-18] well illustrated the success of application of this approach for physical hazards and process safety issues. For instance, models were developed for the thermal decomposition properties of nitro compounds [19-21] or organic peroxides [22-27]. Unfortunately, no QSPR model dedicated to self-reactive substances has been developed to date, probably due to the lack of experimental data.

Indeed, different families of compounds can exhibit self-reactive behaviors, such as azo or diazo compounds, nitro or nitroso compounds, nitrate esters or hydrazine derivatives [2]. But, despite this diversity of compounds, fully documented experimental studies are mainly focused on a limited number of compounds in the literature [6, 28]. For instance, among azo compounds, existing studies on thermal hazards focused on 2,2-azobis(2-methylbutyronitrile) (AMBN) [29], 2,2'-azobis(2,4-dimethylvaleronitrile) (ABVN) [30, 31] and 2,2'-azobis(isobutyronitrile) (AIBN) [1, 32-36].

Considering other types of (non-experimental) estimation approaches, the CHETAH software [37] (developed by ASTM) could be cited but it proposes an estimation of a so-called "Maximum Heat of Decomposition", only suitable for preliminary identification of possible thermal reactivity issues and not as alternative to experimental heats of decomposition. The estimation of this Maximum Heat of Decomposition is based on an extension of the Benson thermochemical group contribution approach and relies on the assumption that CHON compounds decompose in CO_2, H_2O, N_2, CH_4, C, H_2 and O_2, as expected in a complete reaction [38]. So, the calculated Maximum Heat of Decomposition should give a conservative (i.e. higher) value compared to the heats of decomposition obtained in calorimetry measurements.

Consequently, the aim of this paper was to develop the first QSPR models dedicated to the prediction of the heat of decomposition of potential self-reactive substances. Considering the absence of data

available in literature for a sufficiently large diversity of compounds, a new experimental database was consolidated from a dedicated experimental campaign to ensure that all data were obtained in robust and homogenous experimental conditions. This new database allowed the development of first QSPR models following a full validation process including both internal and external validation techniques.

2. Experimental details

The heats of decomposition were measured for 50 solid samples of technical pure substances, supplied and selected by Bayer AG as representing industrial interest. Measurements were performed with a power compensated DSC by Perkin Elmer in stainless-steel high-pressure capsules (up to 150 bar). Capsules consist of a sample container as lower part and a lid with a hole as pressure relief. Both are screwed together and sealed by a gold-plated copper seal. Samples were filled in the sample container under normal atmospheric conditions. Sample mass varied between 3 mg and 7 mg. A heating rate of 5 K/min was applied. As decomposition range of substances varies, the temperature range for the individual measurement was modified to record the decomposition properly. The lowest start temperature was 50 °C, and the highest final temperature was 460 °C. The baseline was determined by a second heating step of the capsule, with the already decomposed sample in it. Measurements were performed twice for each sample, and the final heats of decomposition ($-\Delta H_{dec}$) were defined as the average value. The average deviation between the two runs is 6% with 82% of the observed deviations lower than 10%. One sample (n°8) revealed too large deviation between the two runs (28%) and was excluded. The complete experimental database is available in the Supporting Information (Table S1).

These substances contain chemical groups associated with possible explosive or self-reactive properties like nitro, nitroso, cyano, azo groups or strained rings (Table 1). Three samples (n° 5, 42 and 49) corresponded to salts or hydrated structures and were excluded of the database for sake of structural homogeneity. Two samples (n°31 and 47) concerned were related to the same molecule, methyl-5-chloro-4-nitrothiophene-2-carboxylate. For this reason, only sample n°31 was kept and the retained value of heat of decomposition was the average of the four runs (carried out on both samples n°31 and n°47).
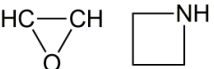
Structural feature	Chemical groups	Examples
Unsaturations	-C=C- ; -C≡C ; -C≡N	Acetylenes, olefins, cyanates
N- and C-Metal atomic pairs	C-Metal ; N-Metal	Organo-lithium compounds
Contiguous N and O atoms	-O-O- ; -N=N	(Di)azo compounds, azides, peroxides
Adjacent N-O atoms	$-NO_2$; -ONO	Nitro compounds, nitrites
N- and O-halogen atomic pairs	N-halogen ; O-halogen	Chlorate, fluoramines
S- and P-O atomic pairs	S=O ; P-O	Sulphonyl halides, phosphites
Strained rings		Epoxides, aziridines

Table 1 – Examples of chemical groups associated to explosive and self-reactive properties

It may be noted that despite this data curation and its small size, the database remains relatively diverse from a chemical point of view with very diverse heterocycles, aliphatic and aromatic nitro

groups, S-based fragments, halogenated structures. If some fragments like NO_2 or =O are found in respectively 34 and 26 compounds of the database, others like C≡N or S=N are less frequent in the database. Moreover, the compounds included in the dataset are often polyfunctional, i.e. presenting different types of functional groups. As a comparison, the databases used for previous QSPR models for the heats of decomposition of more restricted families of compounds, namely organic peroxides [22] and nitroaromatic compounds [39], were constituted of 38 and 77 molecules, respectively.

Finally, the database used in this study consisted of 45 potential self-reactive compounds with a range of decomposition heat values between 61 kJ/mol and 831 kJ/mol. Their structures and $-\Delta H_{dec}$ values (in kJ/mol) are presented in Table 2. The database was then divided into two parts: a training set, containing 34 molecules, and a validation set consisting of the 11 remaining molecules. This partition was performed by a property-ranked approach. The samples were classified by increasing order of heat of decomposition and the ones of the validation set were regularly selected (e.g. 2nd, 6th, 10th, etc.) to represent 1/4 of the dataset. This partitioning method enabled both sets to have similar property distributions in both the training and the validation set (as shown in Figure 1). The similarity of the chemical diversity in both sets was also checked to ensure that the validation set covered at best the chemical diversity of the domain of applicability of the model, i.e. of the molecules in the training set.

Table 2 - Heats of decomposition (kJ/mol) of potential self-reactive substances measured by DSC vs. predicted by the new QSPR models

ID	compounds	- ΔH_{dec} (kJ/mol)		
		exp.	Eq. 3	Eq. 4
Training set				
mol02	N-(dimethyl-lambda4-sulfanylidene)benzenesulfonamide	486.2	460.1	461.8
mol03	1,3-Bis(3,4-dichlorophenyl)-2lambda4-diazathia-1,2-diene	337.4	346.3	344.7
mol04	1-Methyl-5-nitropyridin-2(1H)-one	303.3	356.7	356.7
mol06	4-[(1E)-2-Nitroprop-1-en-1-yl]phenol	243.0	252.7	250.1
mol07	4-Chloro-3-(dimethylnitroryl)benzoic acid	274.0	224.5	227.2
mol09	4-[(Dimethylnitroryl)methyl]-2,6-dimethylphenol	61.0	62.7	66.9
mol10	1-Chloro-4-[(Z)-(4-chlorophenyl)-NNO-azoxy]benzene	371.9	290.2	293.4
mol12	N,N-Diethyl-4-[(E)-(4-nitrophenyl)diazenyl]aniline	458.1	375.6	375.8
mol15	2-[(E)-(4-Chlorophenyl)diazenyl]-1H-imidazole	216.5	275.3	273.0
mol16	2-[(E)-(2,3-Dimethylphenyl)diazenyl]-4,5-dihydro-1,3-thiazole	151.9	194.2	193.3
mol17	(3E)-2-Benzoyl-3-(benzoylimino)-4,5-dimethyl-2,3-dihydro-1,2,3-triazol-3-ium-1ide	157.1	140.4	141.8
mol18	2-[(E)-(3-Chlorophenyl)diazenyl]-1H-imidazole	223.0	269.5	267.2
mol20	2,3-Dimethyl-2,3-dinitrobutane	487.2	507.2	508.1
mol22	2-Nitro-1-(4-nitrophenyl)propane-1,3-diole	422.6	524.7	524.9
mol24	3-Nitro-2H-pyrido[1,2-a]pyrimidin-2,4(3H)-dione	326.9	341.5	341.4
mol25	(5-Nitro-2-phenyl-1,3-dioxan-5-yl)methanol	423.3	380.6	380.9
mol26	1,2,6-Trimethyl-3-nitropyridin-4(1H)-one	359.3	359.4	359.5
mol27	1,2-Dichloro-3,4-dinitrocyclohexane	831.0	815.7	817.3
mol28	Methyl-5-amino-2-nitrocyclohexanecarboxylate	362.7	335.4	333.7
mol29	5-Nitro-2,6-dioxo-1,2,3,6-tetrahydropyrimidine-4-carboxylic acid	239.1	317.7	317.3
mol30	2,6-Dimethyl-3-nitro-1-(2,4,6-trichloro-3-nitrophenyl)pyridin-4(1H)-one	788.2	695.6	696.0
mol31	Methyl-5-chloro-4-nitrothiophene-2-carboxylate	575.8	568.3	568.2
mol32	1-Nitro-4-[(E)-2-nitrobut-1-en-1-yl]benzene	655.3	595.0	596.6
mol34	(3-Methoxy-4-nitro-1H-pyrazol-1-yl)acetic acid	442.6	339.4	339.3
mol35	5-[(3,5-Dimethyl-4-nitro-1H-pyrazol-1-yl)methyl]-2-furancarboxylic acid	314.9	357.0	357.0

mol36	2,2,2-Trichloro-1-(1-methyl-4-nitro-1H-pyrrol-2-yl)ethanone	431.6	503.3	502.6
mol38	Methyl-4-nitro-1H-pyrazol-5-carboxylate	363.1	325.2	324.9
mol39	3-Isobutyl-1-methyl-4-nitro-1H-pyrazol-5-amine	292.4	380.6	380.8
mol40	4-Hydroxy-1-methyl-3-nitroquinolin-2(1H)-one	296.7	286.6	285.2
mol41	1-(1,1-Dioxidotetrahydrothiophen-3-yl)-3,5-dimethyl-4-nitroso-1H-pyrazole	203.8	281.9	281.9
mol43	6-Imino-1-methyl-5-nitroso-1,6-dihydopyrimidin-2,4-diol	222.6	114.0	112.4
mol45	5-Chloro-6-nitro-12H-quinolino[2,1-b]Quinazolin-12-one	398.8	459.4	458.2
mol46	5-Nitro-3,4'-bipyridin-6(1H)-one	300.5	289.9	288.5
mol50	6-Hydroxy-5-nitropyrimidin-4(3H)-one	331.6	326.8	326.6

Validation set

| mol01	1-Nitro-4-(sulfinylamino)benzene	392.6	384.4	384.7
mol11	2-Chloro-4[(E)-phenyldiazenyl]phenol	232.8	210.2	206.6
mol13	(E)-N,N-Dimethyl-2-phenyldiazenecarbothioamide	129.3	170.0	168.9
mol14	4-Amino-6-tert-butyl-3-isopropyl-3,4-dihydro-1,2,4-triazine-5-ol	212.1	158.3	157.1
mol19	7-Nitro-1,2,4-benzotriazin-3(4H)-one-1-oxide	302.4	270.7	273.0
mol21	1-Methyl-4-nitro-1H-pyrazole-5-carboxylic acid	360.4	327.9	327.7
mol23	6-Amino-5-[1-(1,3-benzodioxol-5-yl)-2-nitropropyl]-1,3-dimethylpyrimidin-2,4(1H,3H)-dione	331.2	297.3	296.0
mol33	3-Isobutyl-1-methyl-4-nitro-1H-pyrazole-5-carboxylic acid	280.2	362.3	362.4
mol37	2,2,2-Trichloro-1-[1-(cyclopropylmethyl)-4-nitro-1H-pyrrol-2-yl]ethanone	621.9	506.9	506.2
mol44	2,6-Dimethyl-3-nitro-1-(2,4,5-trichlorophenyl)pyridin-4(1H)-one	484.4	499.8	499.0
mol48	[(5-Methyl-4-nitro-2-thienyl)methylene]malononitrile	423.7	452.6	453.5
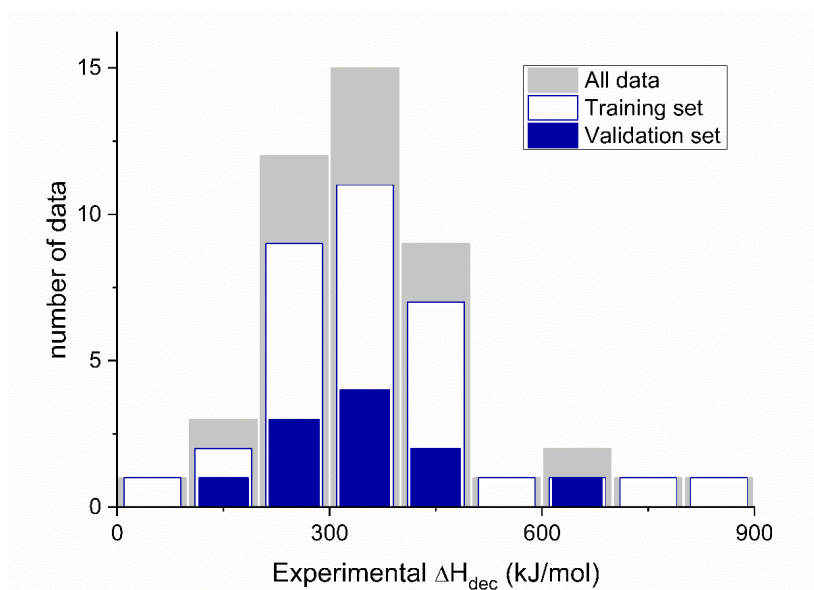
Figure 1 – Distribution of heats of decomposition in the datasets

3. Computational details

3.1. Descriptor calculations

The molecular structures of the 45 potential self-reactive substances of the database were calculated using the density functional theory (DFT) as implemented in the Gaussian16 package [40]. Geometry optimizations were conducted at a B3LYP/6-31+G(d,p) level. Vibrational analyzes were performed with the same functional and basis set. All structures were checked to confirm that no imaginary frequency was calculated. Geometric structures were then encoded into mol files using Openbabel toolbox [41] to be uploaded into the Dragon software (version 7.0.10) [42] that calculated 5255 molecular descriptors. This large diversity of descriptors includes constitutional ones (related to the presence/number of atoms or molecular features in the molecule), topological indices (encoding the two-dimensional molecular structure and providing information about molecular shape and connectivity) and three-dimensional descriptors (including geometric and electronic descriptors, like solvent accessible surface areas and partial charges).

For use in simpler models (without quantum-chemical calculations), the structure of each compound was also encoded into SMILES notation using ChemDraw Ultra 12.0 [43]. Based on these SMILES, Dragon software calculated 3850 two-dimensional molecular descriptors.

3.2. Development of the models

QSARINS software (version 2.2.4) [44, 45] was used for the development of QSPR models. A preliminary filtration step consisted in excluding constant or near-constant descriptors (>80%) and highly inter-correlated ones (> 95%).

Then, multi-linear regression (MLR) models were derived based on the training set of 34 compounds using Ordinary Least Squares (OLS) regression. Descriptor selection was performed using a genetic algorithm (GA). The GA-MLR algorithm was used to search for models containing up to 6 descriptors, assuming that higher-dimension models would be over-parameterized considering the number of

molecules in the training set. A population of 200 models evolved for 2000 generations at each size (i.e. number of descriptors in the model) with a mutation rate of 20%. The leave-one-out (LOO) cross-validation correlation coefficient (Q^2_{LOO}), which will be presented in the next section, was used as the fitness function, since it provides an indicator of the model robustness. The QUIK rule (for Q Under Influence of K) [46] was used to avoid any collinearity between descriptors in models with a threshold of 0.05. Finally, the 20 best models are provided at each size.

3.3.   Validation of the models

All models were validated by both internal and external validation methods. Based on the predictions in the training set, the goodness-of-fit of the model was evaluated based on the determination coefficient (R^2), the mean absolute error (MAE) and the root mean square error (RMSE) between predicted and experimental values. Robustness was assessed by Leave-one-out (LOO) and leave-many-out (LMO) cross-validations. LMO cross-validation was based on 2000 iterations with a 30% partition level. The obtained correlation coefficients Q^2_{LOO} and Q^2_{LMO} may be close to R^2 and close to each other in the case of robust models. To avoid possible chance correlations, a Y-scrambling test [47] was also performed with 2000 iterations of random permutations of the property values within the training set. The average correlation issued from Y-scrambling (R^2_{YS}) was expected to be low.

At last, the predictive power of the model was evaluated by an external validation. The model was applied to the 11 molecules of the validation set, not used for its development. The coefficient of determination R^2_{EXT}, the mean absolute error MAE_{EXT} and the root mean square error $RMSE_{EXT}$ were calculated. A series of external validation metrics were also used: Q^2_{F1} [48], Q^2_{F2} [49], Q^2_{F3} [50], CCC[51], $\overline{r^2_m}$ and Δr^2_m [52]. They were compared to the thresholds proposed by Chirico et al. [51]: $R^2_{EXT} > 0.70$, $Q^2_{Fn} > 0.70$, CCC > 0.85, $\overline{r^2_m} > 0.65$, $\Delta r^2_m < 0.20$.

3.4.   Applicability domain

QSPR are correlation models. Their scope and limitations must be considered. The applicability domain (AD) of the model was analyzed using the leverage approach [53]. The leverage (h) measures the distance of a compound from the rest of the training set in the chemical space (represented by the descriptors).

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

where h_i is the leverage value for compound I, x_i corresponds to its descriptor values and X to the descriptor matrix of the complete training set.

Prediction should not be considered to be reliable if its leverage value exceeds a critical value h^* defined according to Eq. 2.

$$h^* = 3(p+1)/n \qquad (2)$$

where p is the number of descriptors in the model and n is the number of compounds in the training set.

A graphical representation of the AD allowing good identification of outliers is the Williams plot that plots the leverage value vs. standardized residues. In the training set, compounds with $h > h^*$ strongly

influence the statistical parameters of the model. For the test set (or any future prediction), compounds with $h > h^*$ may be considered out of the AD of the model and their predicted values may be not reliable. Compounds presenting standard residuals higher than ± 3 standardized residuals were also considered as statistical response outliers.

4. Results

4.1. Model based on 3D molecular structures

A first model was developed using all the descriptors available in Dragon, i.e. including three-dimensional descriptors. The 5255 molecular descriptors calculated from the 3D molecular structures optimized at B3LYP/6-31+G(d,p) level were loaded into QSARINS. The preliminary filtering of descriptors performed by QSARINS excluded 4008 (constant, nearly constant or intercorrelated) descriptors. The GA-MLR method was then applied on the 34 compounds of the training set using the 1247 remaining descriptors.

Among the models proposed by the GA-MLR algorithm, the one offering the best compromise between correlation, level of parameterization and relevance in terms of chemical interpretation is presented in Eq. 3 and more detailed in Supporting Information (Table S2).

$$-\Delta H_{dec}(kJ/mol) = 5185.7 \; X5Av + 106.3 \; O\text{-}061 - 65.0 \; CATS2D\_02\_DL + 113.3 \; B01[C\text{-}Cl] + 38.9 \quad (3)$$

where X5Av is the average valence connectivity index of order 5, O-061 is the presence of O atoms presenting a delocalized bond, CATS2D_02_DL is the CATS2D Donor-Lipophilic descriptor at lag 02 and B01[C-Cl] is the presence/absence of C and Cl at topological distance 1, i.e. C-Cl bonds.

Despite the small size and the chemical diversity of the dataset, a significant correlation was obtained for the 34 compounds of the training set with $R^2 = 0.88$ and a MAE of 46 kJ/mol (or 14%), as shown in Figure 2. Its robustness was checked by leave-one-out and leave-many-out cross validations ($Q^2_{LOO} = 0.84$; $Q^2_{LMO} = 0.83$) and the Y-scrambling test demonstrated it was not issued from a chance correlation with $R^2_{YS} = 0.12$ (Table 3).
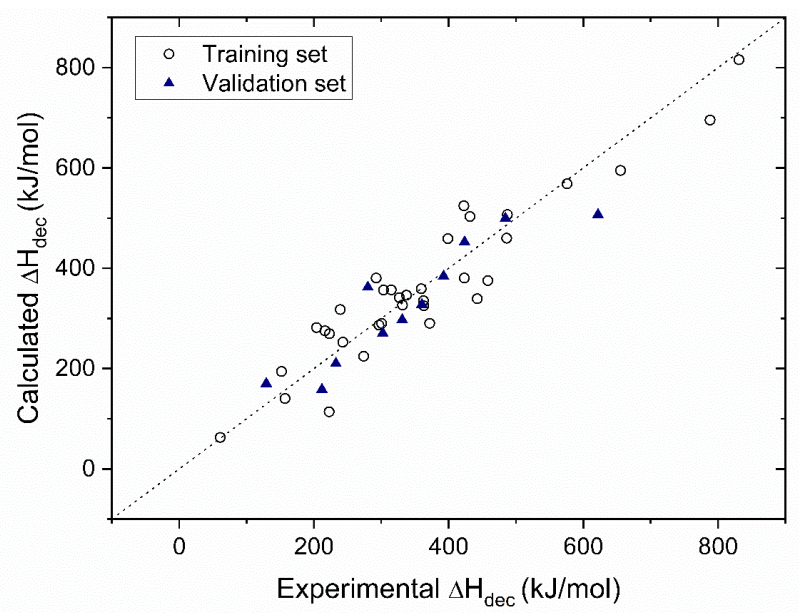

Figure 2 - Experimental vs. calculated heats of decomposition (kJ/mol) for Eq. 3

10

Table 3 – Performances of the QSPR models developed for the prediction of heats of decomposition of potential self-reactive substances based on 3D molecular structures (Eq. 3) and SMILES codes (Eq. 4)

	Eq. 3	Eq. 4
R^2	0.882	0.883
MAE (kJ/mol)	46	45
MAE (%)	14	14
Q^2_{LOO}	0.840	0.843
Q^2_{LMO}	0.825	0.831
R^2_{YS}	0.122	0.122
R^2_{EXT}	0.852	0.851
MAE_{EXT} (kJ/mol)	42	43
MAE_{EXT} (%)	14	14
Q^2_{F1}	0.847	0.845
Q^2_{F2}	0.844	0.842
Q^2_{F3}	0.901	0.899
CCC_{EXT}	0.914	0.913
$<r_m^2>_{EXT}$	0.789	0.787
$\Delta r_m^2{}_{EXT}$	0.106	0.097

The external validation performed on the validation set of 11 molecules revealed similar statistics than in the training set with R^2_{EXT} = 0.85 and MAE_{EXT} = 42 kJ/mol = 14%. Moreover, all Chirico criteria were satisfied with Q^2_{F1} = 0.85, Q^2_{F2} = 0.84 and Q^2_{F3} = 0.90 (i.e. > 0.70), CCC = 0.91 (i.e. > 0.85), $\overline{r^2_m}$ = 0.79 (i.e. > 0.65) and Δr^2_m = 0.11 (i.e. < 0.20).

The analysis of the Williams plot (in Figure 3) evidenced one molecule of the training set, *N*-(dimethyl-λ^4-sulfanylidene)-benzene sulfonamide (mol02), presenting a *h* value higher than the threshold one (0.639 > *h** = 0.441). Nevertheless, the standard residual is reasonable for this molecule. The chemical specificity of this compound could be the presence of the two sulfur-based chemical groups: -S(=O)$_2$- and -S=N<. To evaluate the influence of this molecule on the fitting of the model, it has been excluded from the database and the model Eq. 3 has been fitted again. Similar performances were found both in the training set (with R^2 = 0.89 and MAE = 43 kJ/mol) and in the validation set (R^2_{EXT} = 0.85 and MAE_{EXT} = 41 kJ/mol). So, mol02 didn't degraded Eq. 3 that remained relevant.
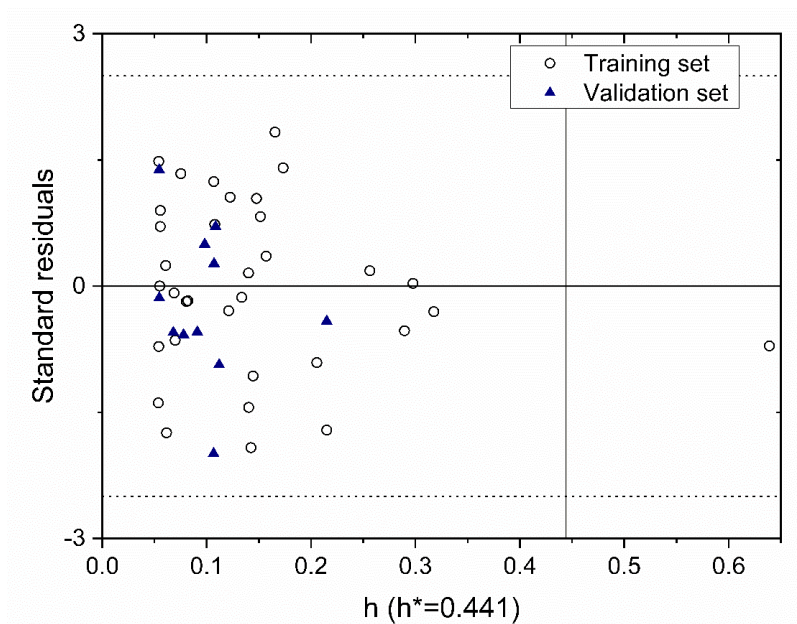
Figure 3 – Williams plot illustrating the applicability domain of Eq. 3

Considering the chemical interpretation of the descriptors in the model, X5Av is the average valence connectivity index of order 5. Average valence connectivity indices are connectivity indices accounting for the presence of heteroatoms, double and triple bonds in the molecule [54, 55]. So, X5Av may relate to the presence of five membered fragments with heteroatoms or double/triple bonds in the molecule. Not surprisingly the highest X5Av value in the dataset was found for mol02 that presents a $-S(=O)_2-N=S(CH_2)_2$ fragment that may greatly contribute in this descriptor. According to the positive coefficient in the equation, such fragments seem to increase the exothermic heat of decomposition of the compound. This agrees with the identification of S=O groups among the chemical groups indicating possible self-reactive properties in the UN Manual of Tests and Criteria [7].

O-061 is the most interpretable descriptor in the model. Indeed, this atom-centred fragment represents the occurrence of O atoms presenting a delocalized bond in the molecule [56]. These atoms are included in nitro groups. So, this descriptor is particularly relevant for self-reactive substances since nitro groups are among the chemical groups associated to possible explosive or self-reactive properties [7, 38]. Besides, the associated regression coefficient in the model is positive, indicating that the presence of nitro groups favors higher heats of decomposition.

The interpretation of the presence of the last two descriptors, CATS2D_02_DL and B01[C-Cl], in the model is less straightforward. But they contribute less to the performance of the model than X5Av and O-061 (based on their associated t-test values, see Supporting Information tables S2 and S4).

CATS2D Donor-Lipophilic at lag 02 relates to the presence of donor and lipophilic groups at a topological distance of 2 in the molecular structure. The detailed analysis of its values in the dataset evidenced that it relates to fragments involving OH, NH_2 groups or aromatic cycles involving an unsaturated N atom (see examples in Figure 4). But none of these fragments were identified as unstable. Besides, the associated regression coefficient in the model is negative which would rather indicate that these groups are more associated to a decrease in exothermic heat of decomposition.
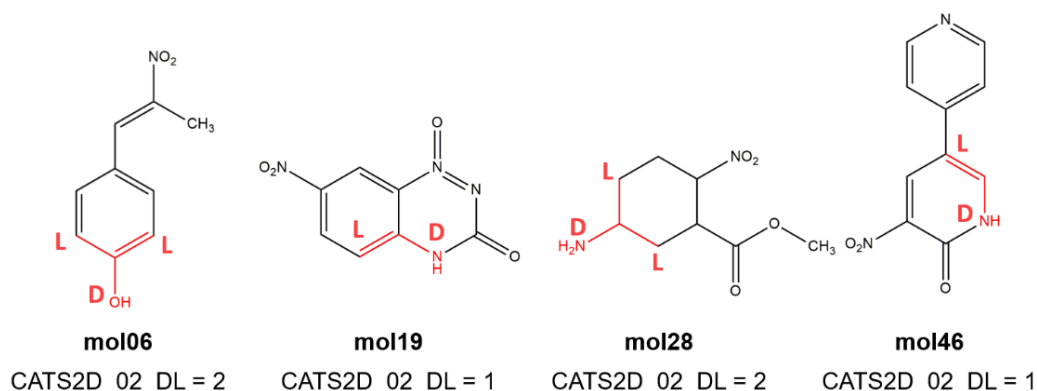
12

Figure 4 – Examples of fragments associated to CATS2D_02_DL.

B01[C-Cl] is related to the presence of C-Cl bonds in the molecule. As the regression coefficient is positive, their presence seems to increase the heat of decomposition, even if C-Cl bonds are not listed in the UN Manual of Tests and Criteria [7] as a chemical group, which indicates explosive or self-reactive properties in organic compounds.

4.2. Model based on SMILES

It may be noticed that none of these 4 descriptors included in this model requires prior 3D structures. So, in the perspective of a simpler application of this model, predictions with Eq. 3 have been done using descriptors calculated from SMILES codes. Finally, only 5 molecules presented deviations in the predictions when the descriptors are extracted by Dragon from quantum-chemical based molecular structures or from SMILES codes (see Table S3). For the global dataset (45 molecules), the mean absolute errors are similar when the calculations were calculated from quantum-chemical structures (MAE = 42.3 kJ/mol) or from SMILES ones (MAE = 42.0 kJ/mol). As highlighted in Supporting Information (Table S3), these deviations can be attributed to the consideration of some nitrogen-nitrogen and nitrogen-oxygen bonds in the molecules, notably for N-O bonds in nitroryl or azoxy groups. As a consequence, different bond orders are found by Dragon from the SMILES codes and from the 3D structures issued from quantum chemical calculations leading to small differences in the calculation of X5Av.

Considering these differences, the model has also been refitted for calculation from SMILES-encoded molecular structure. The refitted model is presented in Eq. 4 and more detailed in Supporting Information (Table S4).

$$-\Delta H_{dec}(kJ/mol) = 5234.5\ X5Av + 107.0\ O\text{-}061 - 66.4\ CATS2D\_02\_DL + 112.3\ B01[C\text{-}Cl] + 36.6 \quad (4)$$

No significant difference is observed in the regression coefficients, and the performance of this SMILES-based model are almost the same as the model regressed for the descriptors calculated from quantum-chemical based molecular structures (Eq. 3), as shown in Table 3.

Accordingly, a similar correlation was observed for the 34 compounds of the training set with $R^2 = 0.88$ and a MAE of 45 kJ/mol (or 14%). The robustness of the refitted model remains good by leave-one-out and leave-many-out cross validations ($Q^2_{LOO} = 0.84$; $Q^2_{LMO} = 0.83$) and the Y-scrambling test demonstrated no chance correlation issues with $R^2_{YS} = 0.12$ (Table 3). The external validation also

performed similar on the validation set of 11 molecules (close to the one observed in the training set) with R^2_{EXT} = 0.85 and MAE_{EXT} = 43 kJ/mol = 14%. Moreover, all Chirico criteria were satisfied with Q^2_{F1} = 0.85, Q^2_{F2} = 0.84 and Q^2_{F3} = 0.90 (i.e. > 0.70), CCC = 0.91 (i.e. > 0.85), $\overline{r^2_m}$ = 0.79 (i.e. > 0.65) and Δr^2_m = 0.10 (i.e. < 0.20).

The GA-MLR method was then used to search for possibly better model based on descriptors calculated from SMILES molecular codes. 3850 two-dimensional molecular descriptors were calculated by Dragon software from the SMILES codes of the 45 potential self-reactive substances investigated in this study. During the preliminary filtering of descriptors performed by QSARINS, 3156 (almost) constant and/or intercorrelated descriptors were excluded. The GA-MLR method was applied on the 34 compounds of the training set using the 694 remaining descriptors. Not surprisingly, the model in Eq. 4 is found as the model representing the best compromises between correlation and level of parameterization.

5. Discussion

Observed errors (about 14 % on the validation set) are slightly higher than observed deviation between two runs (6%). The performances of the models are mainly limited by the small size of the data set compared to its chemical diversity. Nevertheless, these models satisfy all Chirico criteria and already provide promising results for first estimations of heats of decomposition for screening purpose. So, their use for screening purpose should be favored.

A detailed analysis of individual prediction errors was conducted. The distributions of the deviations observed between predicted and experimental values for the different equations are presented in Supporting information (Table S2 and Table S4 for Eq.3 and Eq. 4, respectively). As shown in Figure 5 (for Eq. 3), observed deviations are relatively well distributed (as shown in the Supporting Information).
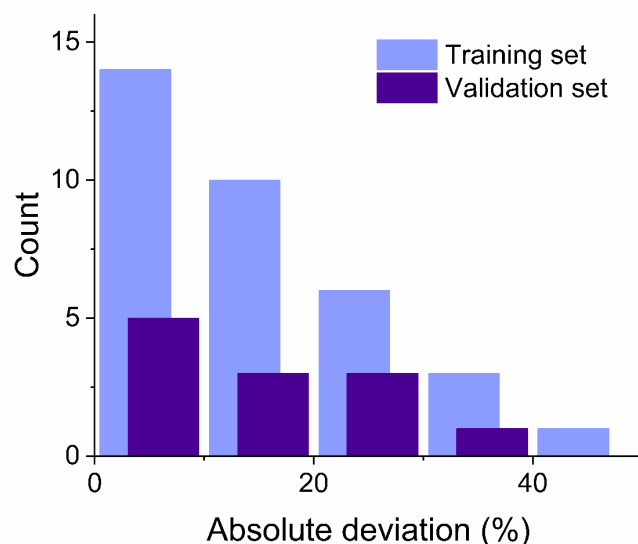

Figure 5 – Distribution of absolute deviations observed between predicted and measured heats of decomposition in the training and in the validation set with model Eq.3

The detailed analysis did not demonstrate any correlation with the differences measured between the two DSC runs conducted during the experimental campaigns. Moreover, associated predictions were all found in the applicability domain of the models as defined by the Williams plots.

As previously mentioned, the dataset used is relatively small compared to the chemical diversity of compounds it contains. In this condition, the quality of prediction can be inhomogeneous inside its applicability domain and molecules presenting some specific molecular features may be poorly predicted. Interestingly, a common molecular feature was evidenced for the compounds presenting the largest deviations (presented in Figure 6). Indeed, both mol41 and mol43 present a terminal N=O fragment, which is not present in any other molecule of the dataset.
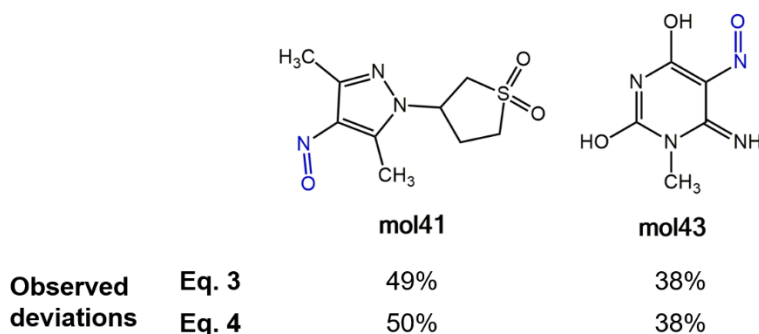

		mol41	mol43
Observed deviations	**Eq. 3**	49%	38%
	Eq. 4	50%	38%

Figure 6 – Compounds presenting the largest deviations (in %) between predictions and experiments.

Due to the relative diversity of molecular structures in the dataset and their polyfunctionality, the further identification of all structural features that could be associated with large deviation is difficult. Nevertheless, the identification of possible structural analogues (presenting the same or similar structural features) can be performed to complement the prediction of a target compound. Indeed, the presence of very similar structural analogues in the training set can improve the confidence of the predictions. Further, the quality of the predictions obtained for structural analogues of high similarity (even out of the training set) can also provide good information on the reliability that can be expected for the predictions on the target compound. Such investigation of structural analogues is anyway recommended when estimating the reliability of QSAR/QSPR prediction for regulatory use and particularly in the framework of REACH [57, 58].

On the 4 descriptors included in these models, two are (electro)-topological indices whereas the other two are constitutional ones. So, they don't account for direct information associated to the 3D or electronic structures of the molecules. Besides, the analysis of their chemical interpretation evidenced their link with the presence of specific groups in the molecule. Moreover, observed performances are similar. So, the SMILES model (Eq. 4) can be considered as the most relevant one.

Finally, these first models represent useful tools to access preliminary estimations of heats of decomposition for screening purpose at early R&D stages, even before synthesis of new substances presenting possible self-reactive behaviors. They can, for instance, be used as screening tools to exclude compounds that could present very high heats of decomposition. The analysis of structural analysis could then be used to improve the confidence of the predictions for the most promising candidates before identifying those that could be selected for synthesis and experimental characterization.

To increase prediction capabilities, improved models should be developed on larger experimental datasets covering a more homogeneous chemical domain (with a better representation of all the structural features involved in the molecules of the dataset). In this study, the samples were selected based upon industrial interest. New experimental campaigns could be conducted including the synthesis of molecules selected more specifically with the aim of a better coverage of the chemical features.

6. Conclusions

In this study, the heats of decomposition of 50 samples of potential self-reactive substances of industrial interest were determined by DSC measurements in homogeneous experimental conditions. After preliminary data curation, the first QSPR models dedicated to potentially self-reactive substances were developed.

Molecular descriptors were calculated by Dragon software either from 3D molecular structures defined from quantum-chemical calculations and from simple SMILES molecular codes. In both cases, good models were obtained and successfully validated in agreement with the OECD principles, including an external validation satisfying the Chirico thresholds on external validation indices.

The model obtained from the descriptors calculated from 3D molecular structures gave promising performances both on the training set (R^2 = 0.88) and in the validation set (R^2_{ext} = 0.85) considering the small size of the database and its chemical diversity. This four-parameter model includes descriptors that can notably be linked to the presence of chemical groups associated to self-reactive or explosive properties.

A simpler model was obtained by focusing on the descriptors calculated from SMILES molecular codes for easier application. Finally, a refitted version of this model was also found offering the same correlation (R^2 = 0.88) and predictive capabilities (R^2_{ext} = 0.85), with only few deviations for molecules presenting molecular features encoded differently into the 3D structures and the SMILES notations (in particular for some nitrogen-nitrogen and nitrogen-oxygen bonds).

If improved models could be developed from a larger database in the future, but these first models already represent a screening tool capable of assessing reactivity hazard information during the development of new chemicals that could exhibit self-reactive behaviors and to anticipate potential process safety issues due to the use of substances presenting potential self-reactive behaviors. To develop better models, larger databases with a better ratio between size and chemical diversity are required. Specific synthesis campaigns specifically designed for this purpose could be of great interest from this perspective.

References

[1] P.-F. Gao, S.-H. Liu, B. Zhang, C.-R. Cao, C.-M. Shu, Complex thermal analysis and runaway reaction of 2,2'-azobis (isobutyronitrile) using DSC, STA, VSP2, and GC/MS, J. Loss Prevent. Proc. Ind., 60 (2019) 87-95.

[2] T. Yoshida, Y. Wada, N. Foster, Safety of Reactive Chemicals and Pyrotechnics, Elsevier, Amsterdam, 1995.

[3] Recommandations on the Transport of Dangerous Goods - Model regulations, 22nd revised Edition, United Nations, 2021.

[4] Globally harmonized system of classification and labelling of chemicals (GHS), ST/SG/AC.10/30/Rev.8, United Nations, New York and Geneva, 2019.

[5] European Commission, Regulation (EC) N°1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) N° 1907/2006, European Commission, 2008.

[6] M. Steensma, P. Schuurman, M. Malow, U. Krause, K.D. Wehrstedt, Evaluation of the validity of the UN SADT H.4 test for solid organic peroxides and self-reactive substances, J. Hazard. Mater., 117 (2005) 89-102.

[7] Manual of Tests and Criteria, 7th Revised Edition, United Nations, New York and Geneva, 2019.

[8] H.G. Fisher, D.D. Goetz, Determination of self-accelerating decomposition temperatures for self-reactive substances, J. Loss Prevent. Proc. Ind., 6 (1993) 183-194.

[9] M. Malow, K.D. Wehrstedt, Prediction of the self-accelerating decomposition temperature (SADT) for liquid organic peroxides from differential scanning calorimetry (DSC) measurements, J. Hazard. Mater., 120 (2005) 21-24.

[10] J. Steinbach, O. Klais, Prediction of Self-Reactive properties from DTA-Measurements, in: HAZARD XIII Symposium, Process Safety-The Future, Manchester, ICHEME Symposium Series No. 141, 1997, pp. 261-272.

[11] Regulation (EC) N° 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), European Commission, 2006.

[12] Guidance Document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models, Organisation for Economic Co-operation and Development (OECD), 2007.

[13] M.T.D. Cronin, A.P. Worth, (Q)SARs for Predicting Effects Relating to Reproductive Toxicity, QSAR Comb. Sci, 27 (2008) 91-100.

[14] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction, Chem. Rev., 110 (2010) 5714-5789.

[15] F.A. Quintero, S.J. Patel, F. Munoz, M.S. Mannan, Review of Existing QSAR/QSPR Models Developed for Properties Used in Hazardous Chemicals Classification System, Ind. Eng. Chem. Res., 51 (2012) 16101-16115.

[16] C. Nieto-Draghi, G. Fayet, B. Creton, X. Rozanska, P. Rotureau, J.-C. De Hemptinne, P. Ungerer, B. Rousseau, C. Adamo, A General Guidebook for the Theoretical Prediction of Physico-Chemical Properties of Chemicals for Regulatory Purposes, Chem. Rev., 115 (2015) 13093-13164.

[17] J.C. Dearden, P. Rotureau, G. Fayet, QSPR prediction of physico-chemical properties for REACH, SAR QSAR Environ. Res., 24 (2013) 545-584.

[18] Z. Jiao, H.U. Escobar-Hernandez, T. Parker, Q. Wang, Review of recent developments of quantitative structure-property relationship models on fire and explosion-related properties, Process Saf. Environ. Protect., 129 (2019) 280-290.

[19] G. Fayet, P. Rotureau, L. Joubert, C. Adamo, Development of a QSPR model for predicting thermal stabilities of nitroaromatic compounds taking into account their decomposition mechanisms, J. Mol. Model., 17 (2011) 2443-2453.

[20] G. Fayet, P. Rotureau, C. Adamo, On the development of QSPR models for regulatory frameworks: The heat of decomposition of nitroaromatics as a test case, J. Loss Prevent. Proc. Ind., 26 (2013) 1100-1105.

[21] Y. Zhang, Y. Pan, J. Jiang, L. Ding, Prediction of thermal stability of some reactive chemicals using the QSPR approach, J. Environ. Chem. Eng., 2 (2014) 868-874.

[22] V. Prana, P. Rotureau, G. Fayet, D. André, S. Hub, P. Vicot, L. Rao, C. Adamo, Prediction of the thermal decomposition of organic peroxides by validated QSPR models, J. Hazard. Mater., 276 (2014) 216-224.

[23] V. Prana, P. Rotureau, D. André, G. Fayet, C. Adamo, Development of Simple QSPR Models for the Prediction of the Heat of Decomposition of Organic Peroxides, Mol. Inform., 36 (2017) 1700024.

[24] Y. Lu, D. Ng, M.S. Mannan, Prediction of the Reactivity Hazards for Organic Peroxides Using the QSPR Approach, Ind. Eng. Chem. Res., 50 (2011) 1515-1522.

[25] Y. Gao, Y. Xue, Z.G. Lu, Z. Wang, Q. Chen, N. Shi, F. Sun, Self-accelerating decomposition temperature and quantitative structure–property relationship of organic peroxides, Process Saf. Environ. Protect., 94 (2015) 322-328.

[26] B. Wang, H. Yi, K. Xu, Q. Wang, Prediction of the self-accelerating decomposition temperature of organic peroxides using QSPR models, J. Therm. Anal. Calori., 128 (2017) 399-406.

[27] Y. Pan, Y. Zhang, J. Jiang, L. Ding, Prediction of the self-accelerating decomposition temperature of organic peroxides using the quantitative structure–property relationship (QSPR) approach, J. Loss Prevent. Proc. Ind., 31 (2014) 41-49.

[28] Y. Yu, K. Hasegawa, Derivation of the self-accelerating decomposition temperature for self-reactive substances using isothermal calorimetry, J. Hazard. Mater., 45 (1996) 193-205.

[29] C.-R. Cao, S.-H. Liu, C.-M. Shu, Reaction simulation of multistage evaluations for AMBN based on DSC experiments, Thermochim. Acta, 661 (2018) 18-26.

[30] S.-H. Liu, Y.-C. Chen, H.-Y. Hou, Thermal runaway hazard studies for ABVN mixed with acids or alkalines by DSC, TAM III, and VSP2, J. Therm. Anal. Calori., 122 (2015) 1107-1116.

[31] S.-H. Liu, Y.-M. Lu, C.-L. Chiang, C.-R. Cao, Determination of the thermal hazard and decomposition behaviors of 2,2'-azobis-(2,4-dimethylvaleronitrile), Process Saf. Environ. Protect., 131 (2019) 55-62.

[32] X.-R. Li, X.-L. Wang, H. Koseki, Study on thermal decomposition characteristics of AIBN, J. Hazard. Mater., 159 (2008) 13-18.

[33] J.P. Van Hook, A.V. Tobolsky, The Thermal Decomposition of 2,2'-Azo-bis-isobutyronitrile, J. Am. Chem. Soc., 80 (1958) 779-782.

[34] M. Malow, K.-D. Wehrstedt, M. Manolov, Thermal decomposition of AIBN Part A: Decomposition in real scale packages and SADT determination, Thermochim. Acta, 621 (2015) 1-5.

[35] B. Roduit, M. Hartmann, P. Folly, A. Sarbach, P. Brodard, R. Baltensperger, Determination of thermal hazard from DSC measurements. Investigation of self-accelerating decomposition temperature (SADT) of AIBN, J. Therm. Anal. Calori., 117 (2014) 1017-1026.

[36] S. Guo, W. Wan, C. Chen, W.H. Chen, Thermal decomposition kinetic evaluation and its thermal hazards prediction of AIBN, J. Therm. Anal. Calori., 113 (2013) 1169-1176.

[37] E.S. Shanley, G.A. Melhem, A review and critique of ASTM CHETAH 4th edition, version 7.2, J. Loss Prevent. Proc. Ind., 13 (2000) 67-68.

[38] T. Yoshida, Safety of Reactive Chemicals, Elsevier, Amsterdam, 1987.

[39] G. Fayet, A. Del Rio, P. Rotureau, L. Joubert, C. Adamo, Predicting the thermal stability of nitroaromatic Compounds using Chemoinformatic Tools, Mol. Inform., 30 (2011) 623-634.

[40] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M.

Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian16, Gaussian Inc., Wallington CT, 2016.

[41] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: An open chemical toolbox, J. Cheminformatics, 3 (2011) 33.

[42] Kode srl, Dragon (Software for Molecular Descriptor Calculation), Pisa, Italy, 2016.

[43] CambridgeSoft Corporation, ChemDraw Ultra 1986-2010.

[44] QSARINS, Varese, Italy, 2018.

[45] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: A new software for the development, analysis, and validation of QSAR MLR models, J. Comput. Chem., 34 (2013) 2121-2132.

[46] R. Todeschini, V. Consonni, A. Maiocchi, The K correlation index: theory development and its application in chemometrics, Chemometr. Intell. Lab., 46 (1999) 13-29.

[47] C. Rücker, G. Rücker, M. Meringer, Y-Randomization and Its Variants in QSPR/QSAR, J. Chem. Inf. Model., 47 (2007) 2345-2357.

[48] A. Tropsha, P. Gramatica, K.V. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, QSAR Comb. Sci, 22 (2003) 69-77.

[49] G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, R. Kühne, External Validation and Prediction Employing the Predictive Squared Correlation Coefficient — Test Set Activity Mean vs Training Set Activity Mean, J. Chem. Inf. Model., 48 (2008) 2140-2145.

[50] V. Consonni, D. Ballabio, R. Todeschini, Comments on the Definition of the Q^2 Parameter for QSAR Validation, J. Chem. Inf. Model., 49 (2009) 1669-1678.

[51] N. Chirico, P. Gramatica, Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection, J. Chem. Inf. Model., 52 (2012) 2044-2058.

[52] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir, Comparative Studies on Some Metrics for External Validation of QSPR Models, J. Chem. Inf. Model., 52 (2012) 396-408.

[53] A.C. Atkinson, Plots, Transformations and regression - An introduction to graphical methods of diagnostic regression analysis, Oxford Science Publications, 1985.

[54] L.B. Kier, L.H. Hall, Derivation and significance of valence molecular connectivity, J. Pharm. Sci., 70 (1981) 583-589.

[55] L.B. Kier, L.H. Hall, General Definition of Valence Delta-Values for Molecular Connectivity, J. Pharm. Sci., 72 (1983) 1170-1173.

[56] A.K. Ghose, G.M. Crippen, Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity, J. Comput. Chem., 7 (1986) 565-577.

[57] Practical guide - How to use alternatives to animal testing to fulfil your information requirements for REACH registration, Version 2.0 ed., European Chemicals Agency (ECHA), 2016.

[58] Practical guide - How to use and report (Q)SARs, Version 3.1 ed., European Chemicals Agency (ECHA), 2016.