Development of validated QSPR models for impact sensitivity of nitroaliphatic compounds

Vinca Prana, Guillaume Fayet, Patricia Rotureau, Carlo Adamo

▶ **To cite this version:**

Vinca Prana, Guillaume Fayet, Patricia Rotureau, Carlo Adamo. Development of validated QSPR models for impact sensitivity of nitroaliphatic compounds. Journal of Hazardous Materials, 2012, 235-236, pp.169-177. 10.1016/j.jhazmat.2012.07.036 . ineris-00961786

HAL Id: ineris-00961786

https://ineris.hal.science/ineris-00961786

Submitted on 20 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Development of validated QSPR models for impact sensitivity of nitroaliphatic compounds

Vinca Prana[a,b], Guillaume Fayet[b], Patricia Rotureau[b,*] and Carlo Adamo[a]

[a] *Laboratoire d'Electrochimie, Chimie des Interfaces et Modélisation pour l'Energie, CNRS UMR-7575, Chimie ParisTech, 11 rue P. et M. Curie, 75231 Paris Cedex 05, France*

[b] *Institut National de l'Environnement Industriel et des Risques (INERIS), Parc Technologique Alata, BP2, 60550 Verneuil-en-Halatte, France*

Abstract

The European regulation of chemicals named REACH implies the assessment of a large number of substances based on their hazardous properties. However, the complete characterization of physico-chemical, toxicological and eco-toxicological properties by experimental means is incompatible with the imposed calendar of REACH. Hence, there is a real need in evaluating the capabilities of alternative methods such as Quantitative Structure-Property Relationship (QSPR) models, notably for physico-chemical properties.
In the present work, the molecular structures of 50 nitroaliphatic compounds were correlated with their impact sensitivities ($h_{50\%}$) using such predictive models. More than 400 molecular descriptors (constitutional, topological, geometrical, quantum chemical) were calculated and linear and multi-linear regressions were performed to find accurate quantitative relationships with experimental impact sensitivities. Considering different sets of descriptors, four predictive models were obtained and two of them were selected for their predictive reliability. To our knowledge, these QSPR models for the impact sensitivity of nitroaliphatic compounds are the first ones being rigorously validated (both internally and externally) with defined applicability domains. They hence follow all OECD principles for regulatory acceptability of QSPRs, allowing possible application in REACH.

Keywords: impact sensitivity; QSPR; nitroaliphatic compounds; quantum chemistry; OECD principles; REACH regulation

*Corresponding author: tel.: + 33(0)344556329; fax: + 33(0)344556565
E-mail address: patricia.rotureau@ineris.fr

1. Introduction

The European Union regulation REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) [1], entered into its registration phase in December 2008. It requires the evaluation of physico-chemical, toxicological and eco-toxicological properties for a tremendous number of chemicals. Indeed, more than 143 000 existing substances have been pre-registered by 65 000 companies and are expected to be registered before 2018. Moreover, all new substances require an early examination to identify possible hazards. Taking into account all the constraints imposed by the regulation, the complete experimental characterization of all substances is not realistic (for reasons of time, costs or ethics in case of tests on animals). Thus, the development of alternative (predictive) methods to evaluate the properties of chemicals was recommended in the framework of REACH.

In this context, we developed and validated theoretical models, based on robust computational approaches, which represent a viable alternative to more conventional experiments. In particular, we focused on Quantitative Structure-Property Relationship (QSPR) models, which have been clearly identified in REACH legislation. Besides, models have already been developed for both environmental and physico-chemical properties in the scope of this regulation [2-4].

To support this development, OECD drawn up the 5 following principles for the validation of QSPR models [5]:

1. A defined endpoint (including experimental protocol);
2. An unambiguous algorithm;
3. A defined domain of applicability;
4. Appropriate measures of goodness-of-fit, robustness and predictive power;
5. A mechanistic interpretation, when it's possible.

Following these criteria, four QSPR models were developed in the present study for the impact sensitivity of nitroaliphatic compounds. This property, characterizing the tendency of a material to react under the effect of a mechanical impact, is one of the most important to classify explosible substances according to the European regulation related to the Classification, Labelling and Packaging of substances and mixtures (CLP) [6].

Many QSPR models have been already developed for this property and, some of them, for nitroaliphatic compounds. The first one was published by Kamlet [7] in 1976 for trinitro-aliphatic compounds and consisted in a simple linear equation based on the oxygen balance that characterizes the oxydability of explosives. Later, Mullay developed another linear model using the molecular electronegativity [8,9]. More recently, quantum chemical descriptors have been introduced. For example, Politzer [10,11], Rice [12] and Badders [13] used the midpoint potential that characterizes the electrostatic potential created by the charges of the C and N atoms at the mid-point of C-NO_2 bonds and Zang [14] exhibited correlations with the nitro group charge. In parallel, Keshavarz [15-18] developed multivariate models using constitutional parameters, such as the ratio of the number of particular molecular fragments on the molecular weight. Lai [19] proposed a new model, based on the ones of Keshavarz and Kamlet that included an additional term based on the presence and the number of specific fragments. All these models were characterized by good determination coefficients (R^2) but they were not validated (either with respect to an internal or external set of compounds) and their respective applicability domains were not defined. Few models using neural networks [20-21] with different sets of descriptors were also developed for nitro compounds.

To our knowledge, the more predictive existing models for the impact sensitivities of nitroaliphatic compounds have been developed by Wang [22] using a multi-linear regression (R^2=0.80, Q^2_{LOO}=0.51 and R^2_{EXT}=0.93) and partial least square (PLS - R^2=0.78, Q^2_{LOO}=0.48 and R^2_{EXT}=0.97) based on a series of electrotopological descriptors. As indicated, the determination coefficients for the molecules of the validation set were high but a failure in

robustness can be noticed probably due to the large number of descriptors (13) in models. Moreover, no applicability domains were defined and analyzed.
Consequently, none of these models fits all OECD standards and cannot be used, therefore, within REACH regulation. In order to fill this gap, this paper aimed to develop fully validated QSPR models to predict the impact sensitivity of 50 nitroaliphatic compounds respecting all OECD principles including the determination of their applicability domains. About 400 descriptors were calculated, including quantum chemical ones based on density functional theory (DFT) calculations. These latter were performed to have a better chemical interpretation of the developed models as already demonstrated in previous works for nitroaromatic compounds [23-25].

At first, linear regressions were performed and a first validated model was proposed. Then several multi-linear regressions (MLR) were developed to reach better performances using different sets of descriptors. The first one considered only constitutional descriptors that only need the 2D structures of molecules. The second one was extended to all types of descriptors and the last one focused only on quantum chemical descriptors that could provide better chemical insights. Also all models were validated by a series of internal and external validation methods. To our knowledge, they represent the first models dedicated to energetic properties proposing a complete validation process including cross validations to characterize robustness, Y-randomization techniques to avoid chance correlations and external validation to estimate predictive power of models and taking into account the applicability domain.

2. Computational details

2.1. Experimental dataset

The considered dataset contained 50 experimental impact sensitivity values of nitroaliphatic compounds (see Table S1 in supporting information for structures) that have been extracted from a single reference [26] to ensure that they were obtained using a single protocol (by drop weight impact tests in the work of Storm [26]), as required by the first OECD point. The impact sensitivity, noted $h_{50\%}$, is a quantity which measures (in centimeters) the height from which the fall of a weight of a given mass on a sample causes a reaction with a 50% probability level. Substances with low $h_{50\%}$ values are the most sensitive since they need less energy to react. This impact test requires relatively small quantities of material (ac700 mg) but does not exhibit good resolution or reproducibility. It is then used to separate qualitatively explosives into classes of similar handling hazards [7]. Also, the test of sensitivity to impact is a regulatory test within the frameworks of CLP [6] or Transport of Dangerous Goods (TDG) [27]. To allow an external validation of models, the data set was then divided into a training set, containing two thirds of the molecules of the data set and a validation set constituted by the remaining molecules. This splitting was carried out to guarantee similar distributions of the $h_{50\%}$ values in both sets [28]. In practice, compounds were classified by increasing order of $h_{50\%}$ and one molecule out of three molecules was selected to constitute the validation set. This partition enables both sets to be of sufficient size with similar distributions to allow a robust development and validation of models. Moreover, no bias in representativeness was identified when inspecting the chemical structures of both sets. The so defined training set of 34 compounds was used for the development and the internal validation of models, while the validation set of 16 compounds was considered for external validation to evaluate their predictive power. The repartition of data between training and validation sets as well as the corresponding experimental values are reported in Table 1.

2.2. Molecular structures calculation

A preliminary conformation analysis was carried out with Scigress software [29] using the Conflex algorithm [30,31] and the MM3 force field [32]. This analysis generated automatically all possible conformations of molecules and their energy. The most stable conformer (the one with the smaller energy) of each molecule was used as molecular structure. Then it was optimizated with Gaussian03 software [33] using DFT calculations with the PBE0 [34] functional and the 6-31+G(d,p) basis set. Vibrational frequencies were computed at same level of calculation to ensure that all final stable conformations exhibited no imaginary frequency.

2.3. Molecular descriptors calculation

Each molecular structure was then characterized by a series of descriptors. According to Todeschini and Consonni [35], the "molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment". Different kinds of calculated descriptors can be then considered [36]:

- Constitutional descriptors, such as the number of specific atoms, functional groups or bonds;
- Topological descriptors, which are calculated from atomic connectivity in the molecule and give information about size, composition and branching degree (e.g. Weiner, Balaban or Randic indices);
- Geometric descriptors, related to the 3D-structure, such as distances, angles, molecular volume or surface area;
- Quantum chemical descriptors, like atomic charges, electronic and binding information, molecular orbital energies or reactivity indices.

Most of the descriptors considered in the present paper were calculated using the CodessaPro software [37]. Others were extracted from examination of chemical structures or from the works of Kamlet [7], Mullay [9], Badders [13] and Keshavarz [17]: oxygen balance (OB_{100}), molecular electronegativity (χ_{mol}), the ratio of the number of atoms (C, H, O or N) on molecular weight (n_C/mw, n_H/mw, n_O/mw, n_N/mw) and the number of specific fragments (as n_{CNC}). Moreover, descriptors issuing from conceptual DFT [38,39], already successfully used in QSPR models for the prediction of the heat of decomposition of nitroaromatic compounds [21,23-25], were considered. More details about descriptors are available in [37] or in supporting information (Tables S2 and S3).

2.4. Model building

All models were developed on the training set using the Best Multi Linear Regression (BMLR) approach [36] as implemented in CodessaPro program [37]. This approach gave the most correlated model at each rank (i.e. for each number of descriptors) and the final model was chosen as the best compromise between correlation and number of descriptors as explained in previous works [24,28].

2.5. Performances of models

To evaluate the performances of models, a series of internal and external validations were computed. The goodness of fit was measured by the root mean square error (RMSE) between predicted and experimental values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n - p - 1}} \qquad (1)$$

With $\hat{y}_i$ the calculated value of the property, y_i the experimental one, $\bar{y}$ the mean of experimental values, n the number of molecules and p the number of descriptors.

Correlation was also characterized through the determination coefficient (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (2)$$

Moreover, a Student's *t*-test validation at a confidence level of 95% was computed to confirm the pertinence of each descriptor. The descriptors with the highest *t*-test values are the most relevant in the model.

The Q^2 coefficient issued from both leave-one-out (LOO) and leave-many-out (LMO) cross validations measured the robustness of the model, i.e. the dependence of the fitting on particular molecules or groups of molecules of the training set.

$$Q^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_{i/i} - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (3)$$

With $\hat{y}_{i/i}$ the predicted value of the i^{th} molecule from the model refitted without using this i^{th} molecule.

Robust models are expected to present a low difference between Q^2 and R^2 coefficients. It has to be noticed that cross validation does not measure the predictivity of models because the set of molecules excluded at each step of the cross validation procedure and then used to calculate Q^2 has already been used for the building of the model [40,41].

To ensure that models did not correspond to chance correlations, a Y-scrambling test was realized. Random permutations of experimental property values were performed (500 iterations) and new models were recalculated [42,43]. To evaluate the impact of randomization, average and standard deviation in R^2_{random} coefficients were calculated (R^2_{YS} and SD_{YS}). Low R^2_{YS} values are expected to avoid chance correlation. To go further, Rücker proposed that the difference between R^2 of the original model and R^2_{YS} should be roughly $\geq 2.3\ SD_{YS}$ to ensure a statistical significance at a 1% level and $\geq 3\ SD_{YS}$ at a 0.1% level [43].

The predictive power of models was measured on the validation set based on the R^2_{EXT} coefficient, which measures the correlation between predicted and experimental values for the molecules of the validation set and the root mean standard error of predictions ($RMSE_{EXT}$). Moreover, the Q^2_{EXT} coefficient (Eq. 4) was calculated as proposed by Tropsha [44] and the OECD guidance document [5].

$$Q_{EXT}{}^2 = 1 - \frac{\sum_{i=1}^{n_{ext}} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_{ext}} (y_i - \bar{y}_{TR})^2} \qquad (4)$$

With $\hat{y}_i$ the predicted value of the property, y_i the experimental value of the property, $\bar{y}_{TR}$ the mean experimental value in the training set and n_{ext} the number of molecules in the validation set.

A QSPR model only offers reliable predictions for compounds similar to those belonging to the training data set. In this study, the applicability domain (AD) [45,46], required by the 3rd OECD point, was determined based on the descriptors included in the model. Euclidean distance method available in Ambit discovery software [47] was used with a 95% threshold, i.e. the domain was calculated to contain 95% of the molecules of the training set. Then, the predictivity inside the applicability domain was also calculated based only on the molecules of the validation set that belonged to this domain (R^2_{IN}, $RMSE_{IN}$ and Q^2_{IN}).

3. Results

3.1. Linear regressions

Firstly, linear regressions were performed on the entire data set to evaluate the global performances of descriptors not only with $h_{50\%}$ but also with $\log h_{50\%}$ as commonly proposed in previous existing models [7,9,13,15-17]. The best correlations were obtained for $\log h_{50\%}$. Indeed, the highest correlation considering $h_{50\%}$ reached R^2=0.57 (obtained with the relative number of aromatic bonds) whereas the ten most correlated descriptors, as shown in Table 2, ranged from R^2=0.73 to 0.78 with $\log h_{50\%}$.

To illustrate this point, oxygen balance (OB_{100}), presenting one of the most correlated descriptors with both $h_{50\%}$ and $\log h_{50\%}$ was represented as function of the experimental impact sensitivity in Figure 1. This descriptor was defined by Kamlet [7] as

$$OB_{100} = 100\ (2n_O - n_H - 2n_C - 2n_{COO})/mw \quad (5)$$

and used as a criterion to evaluate explosibility capacity of CHNO compounds [27]. It is similar to the oxygen balance (OB) defined in the UN regulation related to the transport of dangerous goods (TDG) as a criterion for explosive compounds [27].

From the plot of Figure 1, it appears that the logarithmic fitting is significantly better than the linear one (R^2=0.78 and R^2=0.51, respectively). The same behaviour was noticed for almost all the others descriptors when used alone. Therefore, only $\log h_{50\%}$ was considered in the following of the study.

Moreover, the pertinence of including external descriptors that were not calculated using CodessaPro software was confirmed as they represent four out of the ten most correlated descriptors: OB_{100}, n_C/mw, OB and n_H/mw (see Table 2). Also, some quantum descriptors stood out: the ratio of the zero point vibrational energy on the number of atoms and the ratio of the number of occupied electronic levels on the number of atoms.

Finally, regarding the high correlation of OB_{100} with $\log h_{50\%}$, a first linear model (see Table S4).was developed by considering a complete validation method (including external validation):

$$\log h_{50\%} = 1.714 - 0.205\ OB_{100} \quad (6)$$

It has to be noticed that this model is similar to the equation proposed by Kamlet [7] even if this last was only developed from $C(NO_2)_3$ compounds:

$$\log h_{50\%} = 1.753 - 0.233\ OB_{100} \quad (7)$$

If Kamlet's equation presented better performance in terms of correlation than this new model (R^2=0.94 and R^2=0.72, respectively), it was not validated, in particular on an external set of molecules. On the contrary, the new model (Eq. 6) was developed according to all OECD validation principles. It presented good robustness (Q^2_{LOO}=0.69, Q^2_{10CV}=0.69, Q^2_{5CV}=0.70) and it did not result from chance correlation as checked by Y-scrambling (R^2_{YS}=0.03, SD_{YS}=0.04). Concerning predictivity, the model presented particularly high R^2_{EXT} value (0.92) which is surprisingly largely superior to the R^2 one. Such behavior could be related to the small size of the validation set. Nevertheless, the distribution of property values was quite homogeneous in the validation set as shown on Figure 2. Another explanation could be a structural bias between the training and the validation sets. To explore this hypothesis, a series of 10 random partitions has been set up and models were developed for each of them (Table S5). Seven out of the ten randomized models presented same behaviour ($R^2_{EXT}>R^2$). The average values of R^2_{EXT} remained higher than R^2 ones but none of the randomized partition allowed to reach R^2_{EXT}=0.92 of Equation 6. Moreover, when inspecting the chemical structures of validation sets for each 10 randomized partition, no particular molecule was included in the three partitions presenting $R^2_{EXT}>R^2$ (partitions 1,5,7) and absent in the other seven ones. So, no structural feature was identified to explain this behaviour. Furthermore,

high standard deviations were observed for both R^2_{EXT} (0.08) and R^2 (0.04) which can be explained from high experimental uncertainties.
Nevertheless, this linear model presented already quite good performances and could be used for first evaluation of the property from a single constitutional descriptor which can be easily calculated (i.e. without any software). In a next step, multi-linear regressions were computed to achieve models that could improve prediction.

3.2. Multi-linear regressions

Three models with performances summarized in Table 3 were developed with the BMLR method. The first one was built only from constitutional descriptors which do not require the 3D molecular structures (i.e. without any quantum chemical calculations). Then, all descriptors (about 400) were taken into account for determining a second model, while a third one was defined on the basis of only quantum descriptors.

Model with constitutional descriptors

From 66 constitutional descriptors, a three-parameter model (see Table S6) was found to be the best compromise between correlation and number of descriptors among the six equations sorted out by the BMLR method:

$$\log h_{50\%} = -2.53\ n_N/n_{atom} + 0.07\ n_{single} - 0.25\ n_{NO2} + 1.94 \quad (8)$$

with n_N/n_{atom} the relative number of N atoms (t-test= -2.2), n_{single} the number of single bonds (t-test=7.9) and n_{NO2} the number of NO_2 groups (t-test=-8.4). From a chemical point of view, the occurrence of NO_2 can be explained because the primary cleavage of the C-NO_2 bond is known to be the main mechanism of decomposition [48-51]. The relative number of N atoms and the number of single bonds are not so straightforwardly interpretable.
The predicted impact sensitivity using equation 8 as function of experimental values was plotted in Figure 3. The model was characterized by good correlation (R^2=0.88, RMSE=0.17) and robustness ($Q^2_{LOO}=Q^2_{5CV}$=0.85 and Q^2_{10CV}=0.84). The Y-scrambling method validated the model because of low values of R^2 for the models obtained after randomisation (R^2_{YS}=0.09, SD_{YS}=0.07). As seen in Figure 4, the determination coefficient of each new model (R^2random) can be plotted versus the correlation between original and randomized y-values (R^2Yrandom/Yexp). Besides, the criterion of Rücker [43] was respected: R^2 - R^2_{YS} (0.79) is higher than 3 SD_{YS} (0.22).
The predictive power was high (R^2_{EXT}=0.81, $RMSE_{EXT}$=0.22, Q^2_{EXT}=0.81), in particular in the applicability domain (R^2_{IN}=0.78 and $RMSE_{IN}$=0.23) from which only one molecule (Methylene-bis-N,N'-(2,2,2-trinitroacetamide)) was excluded in the validation set.
This model was fully validated and all descriptors involved in equation 8 can be easily calculated from the 2D-structures of the molecules. Moreover, performances from this multivariate analysis were improved compared to the ones of OB_{100}-based linear model (Eq.6) since it presented higher R^2 than R^2_{EXT} values.

Model with all descriptors

In this section, a model for impact sensitivity was developed by taking into account the whole set of 400 calculated descriptors to find out if a better model including chemically interpretable descriptors could be obtained. A model with four descriptors was derived (see Figure 5 and Table S7):

$$\log h_{50\%} = -0.018\ OB + 4.07\ P_{Qmax\text{-}Qmin} + 28.5\ Q^2_{NO2,max} + 4.80\ N_{O,max} - 0.438 \quad (9)$$

with OB the oxygen balance as defined in the TDG regulation [27], $P_{Qmax\text{-}Qmin}$ the polarity parameter [52,28] defined by the difference between the maximum and minimum charges in the molecule (t-test=7.1), $Q^2_{NO2,max}$ the squared of the maximum NO_2 charges obtained by

natural population analysis [53] (t-test=5.4) and $N_{O,max}$ the maximum nucleophilic reactivity index for a O atom (t-test=2.9).
The first observation is that the OB descriptor which has already been exhibited as one of the most correlated descriptors in linear analysis (see Table 2) is the most important descriptor in the equation (t-test=-14.8). The last three ones are quantum chemical descriptors. $N_{O,max}$ and $Q^2_{NO2,max}$ are related to the electronic and reactivity properties of nitro groups, which are critical in the decomposition process of nitro compounds. $N_{O,max}$ was computed from Codessa software whereas $Q^2_{NO2,max}$ was calculated externally, inspired by the works of Zang [14] and Cao [54]. Besides, the electronic properties of nitro groups also strongly influence the last parameter, $P_{Qmax-Qmin}$.
As shown in Figure 5, equation 9 presented good correlation (R^2=0.93, RMSE=0.13) and robustness ($Q^2_{LOO}=Q^2_{5CV}$=0.90 and Q^2_{10CV}=0.89). The model did not result from chance correlation, since the models issuing from Y-randomization presented low correlations (R^2_{YS}=0.12, SD_{YS}=0.08). The validation set allowed performing an external validation that demonstrated its good predictive power: R^2_{EXT}=0.88, $RMSE_{EXT}$=0.19, Q^2_{EXT}=0.86. Only one molecule of the validation set (1,1,1,6,6,6-hexanitro-3-hexene) was out of the applicability domain. Excluding this molecule, the predictive power remained high with R^2_{IN}=0.88 and $RMSE_{IN}$= 0.19.

Model with quantum chemical descriptors

As three out of the four parameters exhibited from previous equation (Eq.9) were quantum chemical descriptors, a new model using only this class of descriptors was developed (see Table S8). To this end, 172 quantum chemical descriptors were employed and a four parameter model was obtained:

$$logh_{50\%} = -1.6\ N_{occ}/n_{atoms} - 1.4 Q_{O,min} - 0.013\ HDSA1 - 16.5\ BO_{N,max} + 31 \qquad (10)$$

with N_{occ}/n_{atoms} the number of occupied electronic levels divided by the number of atoms (t-test=-13.9), $Q_{O,min}$ the minimal net atomic charge for a O atom (t-test=-4.3), HDSA1 the hydrogen bonding donor ability of the molecule (t-test=-5.1) and $BO_{N,max}$ the maximal bond order for a N atom (t-test=-3.9). It can be noticed that a different set of quantum chemical descriptors was selected (compared to Eq.9) which could be explained by the important intercorrelation between most of these descriptors in particular those characterizing the distribution of charges in the molecules. Among these descriptors, oxygen and nitrogen atoms, that constitute the nitro groups, stand out in this model: the first because of its charge, the second for its bond order. These descriptors are related to the presence of the nitro groups in the molecule and its stability.
This model (Eq. 10) presented also high performances in terms of correlation (R^2=0.90, RMSE=0.16), robustness (Q^2_{LOO}=0.87, Q^2_{10CV}=0.87 and Q^2_{5CV}=0.88) and predictivity (R^2_{EXT}=0.73, $RMSE_{EXT}$=0.29, Q^2_{EXT}=0.68) as seen in Figure 6. Moreover Y-scrambling test ensured against chance correlation (R^2_{YS}=0.12 and SD_{YS}=0.08). For this model, the whole validation set was included into the applicability domain and the predictive power was R^2_{IN}=0.73 and $RMSE_{IN}$=0.30. Finally, this model was less powerful than the two first ones and did not improve the mechanism meaning.

4. Discussion

In this paper, four predictive QSPR models were developed and successfully validated according to the OECD principles for the validation of QSAR/QSPR models. A first linear model (Eq. 6) based on oxygen balance presented sufficient performances for use as a first evaluation method. Then, three models (Eqs. 8-10) presenting better performances were obtained using multi-linear regressions. The three models had comparable performances in terms of fitting (with $R^2 \geq 0.88$ and RMSE≤0.17) and robustness ($Q^2_{LOO} \geq 0.85$) as shown on

Table 3. The first model with only constitutional descriptors (Eq. 8) presented already remarkable predictive power (R^2_{IN}=0.78 and $RMSE_{IN}$=0.23) and the advantage to be easy to use for new predictions, since no advanced calculations were needed to compute the descriptors in these models. Even more accurate predictions (R^2_{IN}=0.88 and $RMSE_{IN}$=0.19) were obtained by including quantum chemical descriptors. Besides, the analysis of the only quantum chemical descriptors led to lower reliability in prediction (R^2_{IN}=0.73 and $RMSE_{IN}$=0.30) demonstrating the pertinence of associating both constitutional and quantum chemical descriptors in the same model.

Table 4 allows comparison of performances between the two new models presenting the better performances (Eqs. 8-9) and those extracted from literature. The new MLR models present better correlation than the model of Mullay [8] (R^2=0.82) and the two ones of Keshavarz [15,18] (R^2=0.77 and 0.85 respectively). Only Kamlet's model [7] is better in fitting (R^2=0.94 vs $R^2 \leq 0.93$ in this study) but without internal or external validation and limited to $C(NO_2)_3$ compounds. In order to have a better comparison with our models, the predictivity of the Kamlet's and Mullay's models was computed using the molecules of the data set not used in the definitions of their models (see definition in Table S1). The computed predictivities for both models (R^2_{EXT}=0.74 and 0.70 for the model of Kamlet and Mullay, respectively) were significantly lower than those obtained for the models developed in this study ($R^2_{EXT} \geq 0.81$). It should be noticed that Keshavarz's models [15,18] were not tested for predictivity because they were developed on the same data set.

The two new models were also compared with Wang's models (MLR and PLS) for nitroaliphatic compounds [22], which were externally validated. The goodness-of-fit of the new MLR models (Eqs. 8-9) were higher than those of Wang's models ($R^2 \geq 0.88$ vs. R^2=0.80 and 0.78 for MLR and PLS respectively). Moreover, internal validation exhibited a high robustness for the two new MLR models ($Q^2_{LOO} \geq 0.85$) whereas Wang's models felt (Q^2_{LOO}=0.51 and 0.48 for MLR and PLS, respectively) due to potential over-parameterization (since 13 descriptors are used in the work of Wang vs. 3 and 4 in this study for Eq.8 and 9 respectively). Furthermore, if the predictivity of Wang's models was also remarkable (R^2_{EXT}=0.93 for MLR and 0.97 for PLS) as the one of the best new MLR model (R^2_{EXT}=0.88, for Eq. 9), Wang did not define any applicability domain. Besides, the model using only constitutional descriptors remained attractive since it associated good predictivity (R^2_{EXT}=0.81) with low computer times.

5. Conclusion

A data set of 50 nitroaliphatic compounds was used to correlate their impact sensitivities to their molecular structures characterized by more than 400 descriptors including quantum chemical ones. Two efficient QSPR models (when considering only constitutional descriptors or all types of descriptors) were built in agreement with all OECD validation principles of QSAR/QSPR models for regulatory use:

1) The endpoint is well defined as the impact sensitivity which characterizes one aspect of explosibility property of chemicals. All experimental values, extracted from a unique database [26], were performed based on the same protocol (drop weight impact test).
2) Algorithms are unambiguous: they consist in multi-linear equations with completely defined descriptors, calculated from DFT optimized structures at PBE0/6-31+G(d,p) in the case of the model using quantum chemical descriptors.
3) Models are applicable to nitroaliphatics. The applicability domains were built by interpolation to include 95% of the molecules of the training set based on the descriptors included in the models.
4) The performances of the models were estimated by various validation tests:

-goodness-of-fit was characterized for the molecules of the training set (R^2, RMSE);
-robustness was evaluated by LOO and LMO cross validations (Q^2_{LOO}, Q^2_{10CV}, Q^2_{5CV});
-chance correlation was checked by Y-scrambling tests (R^2_{YS}, SD_{YS});
-predictive power was defined using an external validation set (R^2_{EXT}, $RMSE_{EXT}$, Q^2_{EXT}), taking also into account the applicability domain (R^2_{IN}, $RMSE_{IN}$, Q^2_{IN}).

5) Some hints were provided concerning the mechanistic interpretation of models. Indeed, several descriptors (notably of quantum chemical type) were identified as related to NO_2 groups that are critical in the chemical mechanism involved in the decomposition process of nitro compounds [48-51].

Finally, among the four models developed in this study for nitroaliphatic compounds, the two selected ones are to the best of our knowledge the first QSPR models that can accurately predict the impact sensitivity of energetic compounds following all OECD principles needed for regulatory use. One of them does not need any advanced quantum chemical calculations and is then easy to use and implement whereas the other including quantum chemical descriptors achieved more reliable predictions. These models could now be used as efficient alternative methods to experimental characterization to gather impact sensitivity values for nitroaliphatic compounds for which experimental values are unavailable or even for compounds that are not yet synthesized (as a preliminary screening test).

Acknowledgements
This work was performed using HPC resources from GENCI-CCRT (Grant 2011-t2011086639). The authors are grateful to Dr Enrico Mombelli (INERIS) for precious discussions.

References

[1] EC (2006) Regulation N°1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the registration, evaluation, authorization and restriction of chemicals (REACH). EC, Brussels.
[2] E. Papa, P. Gramatica, QSPR as a support for the EU REACH regulation and rational design of environmentally safer chemicals: PBT identification from molecular structure, Green Chem. 12 (2010) 836-843.
[3] E. Papa, S. Kovarich, P. Gramatica, On the Use of Local and Global QSPRs for the Prediction of Physico-chemical Properties of Polybrominated Diphenyl Ethers, Mol. Inf. 30 (2011) 232-240.
[4] B. Bhhatarai, W. Teetz, T. Liu, T. Öberg, N. Jeliazkova, N. Kochev, O. Pukalov, I. Tetko, S. Kovarich, E. Papa, P. Gramatica, CADASTER QSPR Models for Predictions of Melting and Boiling Points of Perfluorinated Chemicals, Mol. Inf. 30 (2011), 189-203.
[5] Guidance document on the validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] models, OECD, Paris, 2007.
[6] EC (2008) Regulation N°1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labeling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) N°1907/2006. EC, Brussels.
[7] M.J. Kamlet, The relationship of impact sensitivity with structure of organic high explosives. I. Polynitroaliphatic explosives, In Proceedings of the Sixth Symposium (International) on Detonation, Coronado, California, (1976) 312-322.
[8] J. Mullay, Relationships between impact sensitivity and molecular electronegativity, Propellants Explos. Pyrotech. 12 (1987) 60-63.

[9] J. Mullay, Relationships between impact sensitivity and molecular electronic structure, Propellants Explos. Pyrotech. 12 (1987) 121-124.
[10] P. Politzer, J. S. Murray, P. Lane, A relationship between impact sensitivity and the electrostatic potentials at the midpoints of C-NO_2 bonds in nitroaromatics, Chem. Phys. Lett.181 (1991) 78-82.
[11] P. Politzer, J.S. Murray, Relationships between dissociation energies and electrostatic potentials of C-NO_2 bonds: Applications to impact sensitivities, J. Mol. Struct. THEOCHEM 376 (1996) 419-424.
[12] B.M. Rice, J. J. Hare, A quantum mechanical investigation of the relation between impact sensitivity and the charge distribution in energetic molecules, J. Phys. Chem. A 106 (2002) 1770-1783.
[13] N.R. Badders, C. Wei, A.A. Aldeeb, W.J. Rogers, M.S. Mannan, Predicting the impact sensitivities of polynitro compounds using quantum chemical descriptors, J. Energ. Mater. 24 (2006) 17-33.
[14] C. Zhang, Review of the establishment of nitro group charge method and its applications, J. Hazard. Mater. 161 (2009) 21-28.
[15] M.H. Keshavarz, H.R. Pouretedal, Simple empirical method for prediction of impact sensitivity of selected class of explosives, J. Hazard. Mater. 124 (2005) 27-33.
[16] M.H. Keshavarz, M. Jaafari, Investigation of the various structure parameters for predicting impact sensitivity of energetic molecules via artificial neural network, Propellants Explos. Pyrotech. 31 (2006) 216-225.
[17] M.H Keshavarz, H.R. Pouretedal, A. Semnani, Novel correlation for predicting impact sensitivity of nitroheterocyclic energetic molecules, J. Hazard. Mater.141 (2007) 803-807.
[18] M.H. Keshavarz, Prediction of impact sensitivity of nitroaliphatic, nitroaliphatic containing other functional groups and nitrate explosives, J. Hazard. Mater.148 (2007) 648-652.
[19] W.P. Lai, P. Lian, B.Z. Wang, Z.X. Ge, New correlations for predicting impact sensitivities of nitro energetic compounds, J. Energ. Mater. 28 (2010) 45-76.
[20] S.G. Cho, K.T. No, E.M. Goh, J.K. Kim, J.H. Shin, Y.D. Joo, S. Seong, Optimization of neural networks architecture for impact sensitivity of energetic molecules, Bull. Korean Chem. Soc. 26 (2005) 399-408.
[21] H. Nefati, J.M. Cense, J.J. Legendre, Prediction of the impact sensitivity by neural networks, J. Chem. Inform. Comput. Sci. 36 (1996) 804-810.
[22] R. Wang, J.C. Jiang, Y. Pan, H.Y. Cao, Y. Cui, Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices, J. Hazard. Mater.166 (2009) 155-186.
[23] G. Fayet, L. Joubert, P. Rotureau, C. Adamo, On the use of descriptors arising from the conceptual density functional theory for the prediction of chemicals explosibility, Chem. Phys. Lett. 467 (2009) 407-411.
[24] 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.
[25] G. Fayet, A. Del Rio, P. Rotureau, L. Joubert, C. Adamo, Predicting the thermal stability of nitroaromatic compounds using chemoinformatic tools, Mol. Inf. 30 (2011) 623-634.
[26] C.B. Storm, J.R. Stine, J.F. Kramer, Sensitivity relationships in energetic materials. Chemistry and physics of energetic materials, Kluwer Academic Publishers, 1990.
[27] Recommendations on the transport of dangerous goods: Manual of tests and criteria, ST/SG/AC.10/Rev.5 fifth revised edition, United Nations, Geneva/New-York, 2011.

[28] A.R. Katritzky, I. B. Stoyanova-Slavova, D.A. Dobchev, M. Karelson, QSPR modeling of flash points: An update, J. Mol. Graph. 26 (2007) 529-536.
[29] Scigress, version 2.2.0, Fujitsu limited, 2008-2011.
[30] H. Goto, E. Osawa, An efficient algorithm for searching low energy conformers of cyclic and acyclic molecules, J. Chem. Soc., Perkin Trans. 2.(1993) 187-198.
[31] H. Goto, E. Osawa, Corner flapping – a simple and fast algorithm for exhaustive generation of ring conformations, J. Am. Chem. Soc. 111 (1989) 8950-8951.
[32] N.L. Allinger, Y.H. Yuh, J.H. Lii, Molecular mechanics. The MM3 forcefield for hydrocarbons. 1, J. Am. Chem. Soc. 111 (1989) 8551–8566.
[33] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A.Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N.Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V.Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson,H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda,J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H.Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V.Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann,O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski,P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg,V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain,O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B.Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski,B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi,R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng,A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W.Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian03, 2004, Gaussian Inc. http://www.gaussian.com.
[34] C. Adamo, V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, J. Chem. Phys. 110 (1999) 6158-6170.
[35] R. Todeschini, Handbook of molecular descriptors, Wiley-VCH, Weinheim; New York, 2000.
[36] M. Karelson, Molecular Descriptors in QSAR/QSPR, Wiley, New York, 2000.
[37] CodessaPro, 2002, University of Florida.
[38] H. Chermette, Chemical reactivity indexes in density functional theory, J. Comput. Chem. 20 (1999) 129-154.
[39] P. Geerlings, F. De Proft, W. Langenaeker, Conceptual density functional theory, Chem. Rev. 103 (2003) 1793-1873.
[40] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694-701.
[41] A. Tropsha, Best Practices for QSAR Model Development, Validation and Exploitation, Mol. Inf. 29 (2010) 476-488.
[42] F. Lindgren, B. Hansen, W. Karcher, Model validation by permutation tests: applications to variable selection, J. Chemometr. 10 (1996)521-532.
[43] C. Rücker, G. Rücker, M. Meringer, y-Randomization and Its Variants in QSPR/QSAR, J. Chem. Inf. Moldel. 47 (2007) 2345-2357
[44] A. Tropsha, P. Gramatica, V. K. 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.
[45] T.I. Netzeva, A. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J.M. van de Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52, Altern Lab Anim. 33 (2005) 155-173.

[46] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, Altern Lab Anim. 33 (2005) 445-459.
[47] N. Jeliazkova, J. Jaworska, Ambit Discovery, version 1.20, 2007.
[48] M.J.S. Dewar, J.P. Ritchie, J. Alster, Ground-state of molecules.65. Thermolysis of molecules containing NO_2 groups, J. Org. Chem. 50 (1985) 1031-1036.
[49] G.M. Nazin, G.B. Manelis, Thermal-decomposition of aliphatic nitrocompounds, Russ. Chem. Rev. 63 (1994) 313.
[50] G.M. Nazin, G.B. Manelis, G.N. Nechiporenko, F.I. Dubovitsky, Thermal decomposition of some polynitrocompounds in gas phase, Combust. Flame. 12 (1968) 102-106.
[51] V.L. Korolev, T.S. Pivina, A.A. Porollo, T.V. Petukhova, A.B. Sheremetev, V.P. Ivshin, Differentiation of the molecular structures of nitro compounds as the basis for simulation of their thermal destruction processes, Russ. Chem. Rev. 78 (2009) 945-969.
[52] K. Osmialowski, J. Halkiewicz, A. Radecki, R. Kaliszan, Quantum chemical parameters in correlation analysis of gasliquid chromatographic retention indices of amines, J. Chromatogr.A346 (1985) 53-60.
[53] A.E. Reed, L.A. Curtiss, F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint, Chem. Rev. 88 (1988) 899-926.
[54] C.Z. Cao, S. Gao, Two dominant factors influencing the impact sensitivities of nitrobenzenes and saturated nitro compounds, J. Phys. Chem. B 111 (2007) 12399-12402.

Figure 1. Experimental impact sensitivity ($h_{50\%}$ in cm) vs Oxygen Balance (OB_{100}). Linear and logarithmic fittings are also indicated in blue full line and red dotted line respectively

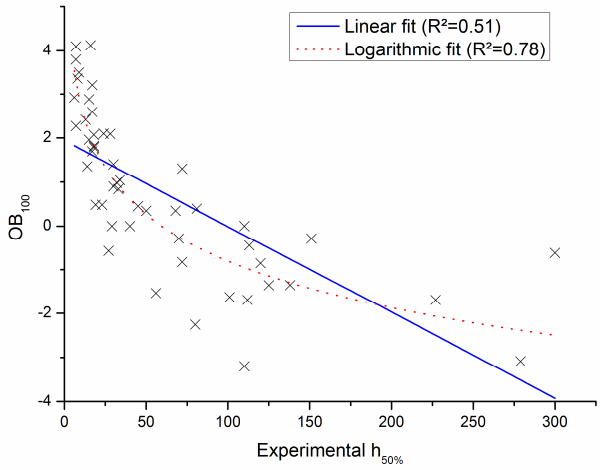

Figure 2. Experimental vs. predicted $\log h_{50\%}$ from the linear model based on Oxygen Balance (OB_{100}) (Eq. 6)

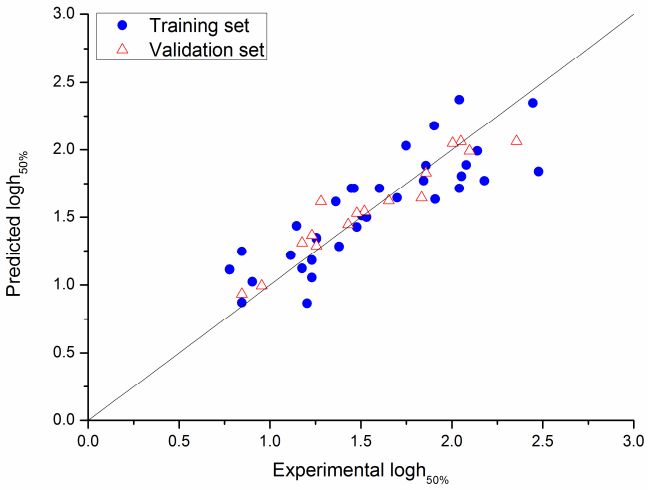

Figure 3. Experimental vs. predicted $\log h_{50\%}$ from the model based on constitutional descriptors (Eq. 8)

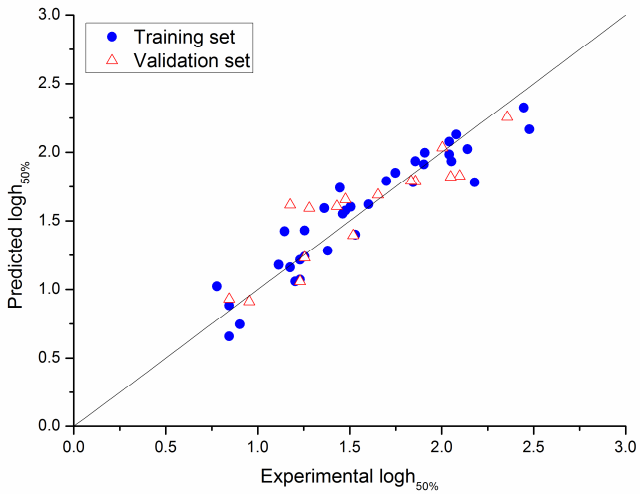

Figure 4. Results of Y-scrambling test in terms of $R^2_{Yrandom/Yexp}$ vs. R^2_{random} for the model based on constitutional descriptors (Eq. 8)

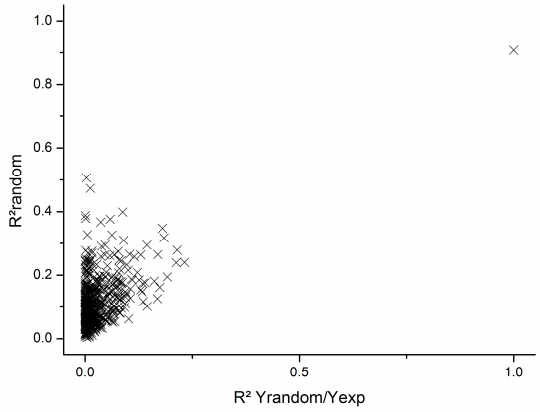

Figure 5. Experimental vs. predicted $logh_{50\%}$ from the model based on all calculated descriptors (Eq. 9)

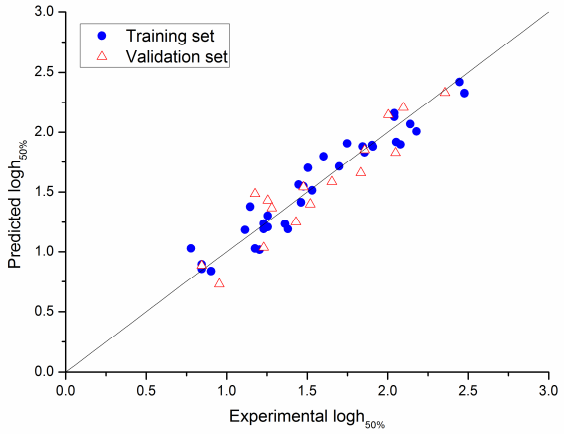

Figure 6. Experimental vs. predicted $logh_{50\%}$ from the model based on quantum chemical descriptors (Eq. 10)

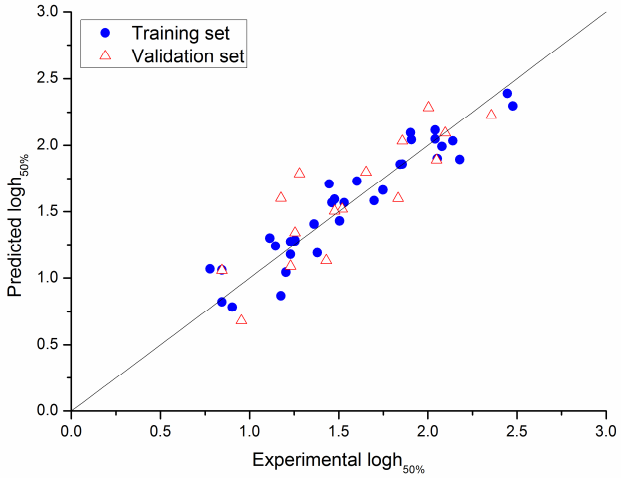

Table 1: **Experimental and predicted values for $h_{50\%}$ (from Eqs. 6, 8, 9, 10)**

	Molecules	**$h_{50\%\ exp}$**	**$logh_{50\%\ exp}$**	**Eq. 6**	**Eq. 8**	**Eq. 9**	**Eq. 10**
	Training Set						
mol1	1,1,1,3,5,5,5-heptanitropentane	8	0.90	1.02	0.75	0.84	0.79
mol2	1,1,1,6,6,6-hexanitro-3-hexyne	7	0.85	1.25	0.88	0.90	0.82
mol3	3,3,4,4-tetranitrohexane	80	1.90	2.18	1.91	1.89	2.10
mol4	2,2,4,4,6,6-hexanitroheptane	29	1.46	1.71	1.55	1.41	1.57
mol5	2,2,4,6,6-pentanitroheptane	56	1.75	2.03	1.85	1.91	1.66
mol6	2,2,2-Trinitroethyl-carbamate	18	1.26	1.35	1.43	1.30	1.28
mol7	2,2-Dinitro-1,3-propane-diol	110	2.04	1.71	1.98	2.16	2.05
mol8	Methyl-2,2,2,-trinitro-ethyl carbonate	28	1.45	1.71	1.74	1.56	1.71
mol9	4,4,4-trinitrobutyramide	40	1.60	1.71	1.62	1.80	1.73
mol10	Bis-(2,2,2trinitroethyl)-carbonate	16	1.20	0.87	1.06	1.02	1.04
mol11	Bis-(trinitroethoxy)-methane	17	1.23	1.05	1.22	1.24	1.27
mol12	N,N'-Bis-(2,2,2-trinitroethyl)-urea	17	1.23	1.18	1.07	1.19	1.18
mol13	Ethyl-2,2,2-trinitro-ethylcarbonate	81	1.91	1.63	2.00	1.88	2.04
mol14	Bis-(trinitroethyl)-oxalate	15	1.18	1.12	1.16	1.03	0.87
mol15	Bis-(trinitroethyl)-oxamide	13	1.11	1.22	1.18	1.18	1.30
mol16	N-trinitroehyl-4,4,4-trinitrobutyramide	18	1.26	1.34	1.24	1.21	1.28
mol17	1,5-bis-(trinitroethyl)-biuret	24	1.38	1.28	1.28	1.19	1.19
mol18	N-(t-butyl)-trinitro-acetamide	110	2.04	2.37	2.08	2.13	2.12
mol19	1,1,1,7,7,7-hexanitroheptanone-4	34	1.53	1.50	1.40	1.51	1.57
mol20	2,2-dinitropropyl-trinitrobutyrate	151	2.18	1.77	1.78	2.01	1.89
mol21	2,2,2-trinitroethyl-4,4-dinitrovalerate	70	1.85	1.77	1.78	1.88	1.86
mol22	Bis-(2,2-dinitropropyl)-carbonate	300	2.48	1.84	2.17	2.33	2.30
mol23	Bis-(trinitropropyl)-urea	23	1.36	1.61	1.59	1.24	1.41
mol24	Bis(trinitroethyl)fumarate	14	1.15	1.44	1.42	1.38	1.24
mol25	Trinitroethyl-bis-(trinitroethoxy)-acetate	6	0.78	1.11	1.02	1.03	1.07
mol26	4,4,4-trinitrobutyric anhydride	30	1.48	1.43	1.58	1.55	1.59
mol27	2,2,2 trinitroethyl-4,4-dinitrohexanoate	138	2.14	1.99	2.02	2.07	2.03
mol28	Nitroisobutyl-4,4,4-trinitrobutyrate	279	2.45	2.35	2.33	2.42	2.39
mol29	Tetrakis-(2,2,2-trinitro-ethyl)-orthocarbonate	7	0.85	0.87	0.66	0.86	1.06
mol30	Methylene-bis-(4,4,4-trinitrobutyramide)	113	2.05	1.80	1.93	1.92	1.90
mol31	Ethylene-bis-(4,4,4-trinitrobutyrate)	120	2.08	1.89	2.13	1.90	1.99
mol32	N,N'-Bis-(2,2-dinitro-propyl)-4,4,4-trinitro-butyramide	72	1.86	1.88	1.94	1.83	1.86
mol33	2,2-dinitropropane-1,3-1,3-diol-bis-(4,4,4-trinitro butyrate)	50	1.70	1.64	1.79	1.71	1.58
mol34	Bis-(2,2,2trinitroethyl)-4,4,6,6,8,8-hexanitro-undecanedioate	32	1.51	1.51	1.60	1.70	1.43
	Validation Set						
mol35	1,1,1,3-tetranitrobutane	33	1.52	1.54	1.39	1.40	1.52
mol36	1,1,1,6,6,6-hexanitro-3-hexene	17	1.23	1.37	1.06	1.04	1.09
mol37	Methylene-bis-N,N'-(2,2,2-trinitroacetamide)	9	0.95	0.99	0.91	0.73	0.68
mol38	5,5,5-trinitropentanone-2	125	2.10	1.99	1.83	2.21	2.09
mol39	N-(2-propyl)-trinitroacetamide	112	2.05	2.06	1.82	1.83	1.89
mol40	2,2,2-Trinitroethyl-4,4,4-trinitrobutyrate	18	1.26	1.29	1.23	1.43	1.34
mol41	Trinitroethyl-2,2-dinitropropylcarbonate	15	1.18	1.31	1.62	1.49	1.60
mol42	Tris-(2,2,2-trinitroethyl)-orthoformate	7	0.85	0.93	0.93	0.88	1.06
mol43	Methylene-bis-(trinitroethyl)-carbamate	27	1.43	1.45	1.61	1.25	1.13

mol44	2,2 dinitropropyl–4,4,4-trinitrobutyramide	72	1.86	1.83	1.79	1.85	2.03
mol45	Bis-(1,1,1-trinitro-2-propyl)-urea	19	1.28	1.61	1.59	1.37	1.79
mol46	Bis-(2,2,2-trinitroethyl)-succinate	30	1.48	1.53	1.66	1.54	1.50
mol47	Bis-(2,2-dinitropropyl)-oxalate	227	2.36	2.06	2.26	2.33	2.23
mol48	N,N'-bis-(3,3,3-trinitro-propyl)-oxamide	45	1.65	1.62	1.69	1.59	1.80
mol49	2,2 dinitrobutyl-4,4,4-trinitrobutyramide	101	2.00	2.05	2.03	2.15	2.29
mol50	Bis-(2,2,2-trinitro-ethyl)-4,4-dinitro-ethyl)-4,4-dinitroheptanedioate	68	1.83	1.64	1.79	1.66	1.60

Table 2: Descriptors presenting the best linear correlations with $logh_{50\%}$ for the whole data set

R^2	**Descriptors**
0.78	Relative number of aromatic bonds
0.78	Oxygen balance (OB_{100}) according to Kamlet [7]
0.77	Ratio of number of C atoms on molecular weight
0.77	Relative number of single bonds
0.76	Oxygen balance (OB) according to regulation [27]
0.76	Ratio of zero point vibrational energy on the number of atoms
0.76	Ratio of number of occupied electronic levels on the number of atoms
0.76	Relative molecular weight
0.75	Relative number of H atoms
0.73	Ratio of number of H atoms on molecular weight

Table 3: Performances of QSPR models developed for $logh_{50\%}$ of nitroaliphatic compounds in this study

	Eq. 6	**Eq. 8**	**Eq. 9**	**Eq. 10**
R^2	0.72	0.88	0.93	0.90
RMSE	0.15	0.17	0.13	0.16
Q^2_{LOO}	0.69	0.85	0.90	0.87
Q^2_{10CV}	0.69	0.84	0.89	0.87
Q^2_{5CV}	0.70	0.85	0.90	0.88
R^2_{YS}	0.03	0.09	0.12	0.12
SD_{YS}	0.04	0.07	0.08	0.08
R^2_{EXT}	0.92	0.81	0.88	0.73
$RMSE_{EXT}$	0.15	0.22	0.19	0.29
Q^2_{EXT}	0.89	0.81	0.86	0.68
R^2_{IN}	0.92	0.78	0.88	0.73
$RMSE_{IN}$	0.15	0.23	0.19	0.30
Q^2_{IN}	0.89	0.78	0.87	0.68

Table 4: **Performances of the two more reliable QSPR models developed for $logh_{50\%}$ of nitroaliphatic compounds in this study and extracted from literature**

Models	N_{mol}/N_{desc}	R^2	RMSE	Q^2_{LOO}	R^2_{EXT}	$RMSE_{EXT}$
Eq. 8 (constitutional descriptors)	50/3	0.88	0.17	0.85	0.81	0.22
Eq. 9 (all descriptors)	50/4	0.93	0.13	0.90	0.88	0.19
Kamlet [7][b]	28/1	0.94	N.A.	N.A.	0.74[a]	N.A.
Mullay [8]	41/1	0.82	N.A.	N.A.	0.70[a]	N.A.
Keshavarz, 2005 [18]	58/5	0.85	0.20	N.A.	N.A	N.A.
Keshavarz, 2007 [15]	58/4	0.77	0.19	N.A.	N.A	N.A.
Wang [22]	43/13	0.80	0.20	0.51	0.93	0.19

[a] determined based on the molecules of the dataset that were not used for the development of models in the original work (see supporting information)

[b] only applicable for $C(NO_2)_3$ compounds