Global and local quantitative structure-property relationship models to predict the impact sensitivity of nitro compounds

Guillaume Fayet, Patricia Rotureau, Vinca Prana, Carlo Adamo

▶ To cite this version:

HAL Id: ineris-00961780

https://hal-ineris.ccsd.cnrs.fr/ineris-00961780

Submitted on 20 Mar 2014

Global and local QSPR models to predict the impact sensitivity
of nitro compounds

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

[a] INERIS, Parc Technologique Alata, BP2, 60550 Verneuil-en-Halatte, France

[b] 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

*Corresponding author: patricia.rotureau@ineris.fr ; tel: +33(0)344616329 ; fax: +33(0)344556565

ABSTRACT

New quantitative structure property relationships (QSPR) were developed to predict accurately the impact sensitivity of nitro compounds from their molecular structures. Such predictive approaches represent good alternative to complete experimental testing in development process or for regulatory issues (e.g. within the European REACH regulation).

To achieve highly predictive models, two approaches were used to explore the whole diversity of nitro compounds included in a data set of 161 molecules. In a first step, local models, dedicated to nitramines, nitroaliphatics and nitroaromatics, were proposed. After that, a global model was developed to be applicable for the whole range of the nitro compounds of the data set.

In both cases, large series of molecular descriptors were calculated from quantum chemically calculated molecular structures and multilinear regressions were computed to correlate them with experimental impact sensitivities. Both global and local models led to high accuracy in predictions for nitramines and nitroaliphatics whereas nitroaromatics revealed more difficult to predict due to their complex decomposition mechanisms.

The proposed models were validated in the perspective of potential regulatory use according to the OECD principles, including internal, external validation and the definition of their applicability domain. So, they could then be used for prediction either separately or in a consensus approach.

Keywords: Quantitative Structure-Property Relationships; Density Functional Theory; Validation; Applicability domain; Impact Sensitivity; Nitro Compounds.

INTRODUCTION

Sensitivity is a critical property of energetic materials as it characterizes the tendency of the material to react under the effect of various stimuli: heat, electric spark, friction or impact. In particular, impact sensitivity, denoted $h_{50\%}$, expresses the sensitivity of materials to a mechanical impact. It is experimentally evaluated by determining the height at which a weight of a defined mass leads to a reaction when falling on a material sample, with a 50% level of probability. It is a standard property of explosive substances not only when evaluating performances but also for safety considerations, since it is part of the regulatory tests within regulatory frameworks, e.g. Globally Harmonized System (GHS) [1] or Transport of Dangerous Goods (TDG) [2].

To evaluate as early as possible potential safety problems when handling energetic materials, notably in development process, the use of computation approaches is of great interest. Computer capabilities are in constant increase and developing accurate predictive methods to estimate the properties of such compounds is now possible. In particular, Quantitative Structure-Property Relationships (QSPR) are increasingly used to predict properties of chemicals [3-10]. This approach consists in correlating an experimental property with descriptors of the molecular structure. Such models are even recommended into the new European regulation related to the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) [11] and rules have been defined by the Organisation for Economic Co-operation and Development (OECD) to clarify their validation for a regulatory use [12]. These rules not only clarify the importance of transparency in the definition of the algorithm, but also in the definition of endpoints (including experimental protocols) and domains of applicability, as well as in the measures of performances of models (e.g. with an external validation to estimate predictivity). Moreover, they encourage their mechanistic interpretation.

Some QSPR models have been already devoted to impact sensitivity in the past. Kamlet [13,14] proposed linear models linking $\log h_{50\%}$ to the oxygen balance for nitro compounds with high correlations but without any external validation. Correlations were also proposed with the charge distribution descriptors [15], e.g. empirical molecular electronegativities [16,17], charge of nitro groups [18,19], electrostatic potential at the midpoint of the $C-NO_2$ bond (V_{mid}) [15,20,21], or the dissociation energy of this bond [20,22,23]. These correlations notably highlighted the critical part of the $C-NO_2$ bond in the decomposition process of nitro compounds [24].

In more recent works, multivariate approaches were preferred. For instance, Nefati [25], Cho [26] and Jun [27] used artificial neural networks (ANN) with various descriptors. Li [28] also proposed models based on the $C-NO_2$ bond dissociation energy and the oxygen balance. A series of multi linear regressions were also proposed by Keshavarz [29-33], based on constitutional descriptors for various nitro compounds, or Afanas'ev [34] and Badders [35], with a larger diversity of descriptors. Most of these models presented remarkable correlations but their predictive powers were in general not evaluated on an external set of molecules.

Only few models were proposed with a transparent validation on an external set. Morrill [36] proposed MLR models based on quantum chemical descriptors issued from semi-empirical calculations with only low performances in prediction (56% in relative error). Keshavarz [37] also proposed a model dedicated to nitramine, nitroaromatic and nitroaliphatic compounds with a root

mean squared deviation in prediction of 23 cm. However, this last model was not checked by internal validation tests (neither cross validation, nor Y-randomization).

In the work of Wang [38], various approaches (MLR, partial least squares and ANN) were used to develop models based on electrotopological descriptors to predict the impact sensitivities of nitroaromatic, nitramine and nitroaliphatic compounds. High predictive powers were exhibited (up to $R^2_{ext}=0.932$, for MLR models) but lack in robustness can be noticed, potentially due to over-parameterization since models are based on up to 16 descriptors. Moreover, the domains of applicability of the models were not considered. Thus, these models do not achieve all OECD principles for their validation.

In this paper, new models were developed with aim to follow all these principles and, then, be applicable in a regulatory framework. In particular, multilinear models were proposed to predict the impact sensitivities of nitroamine, nitroaliphatic and nitroaromatic compounds using a wide variety of descriptors, including quantum chemical descriptors that are expected to allow interpretation in terms of chemical reactivity. These models were validated based on a series of internal and external validation tests. Nitro compounds were first considered by families to obtain local models and then in the same database to achieve a unique global model.

MATERIALS AND METHODS

Experimental data set

This paper aims to correlate the impact sensitivities ($h_{50\%}$) to the chemical structures of nitro compounds. Nevertheless, if the chemical mechanisms involved in the initiation of decomposition by an impact, at a molecular scale, are linked to chemical structures, experimental conditions and physical effects, notably related to granulometry, also influence measurements. Moreover, they are thus not completely reproducible, in particular from an apparatus to another [14].

Therefore, the choice of experimental data is critical since all uncertainties are propagated in the final model during the fitting procedure. So, to ensure at best the compatibility between experimental values, all experimental impact sensitivities were extracted from a single reference [39], collecting the impact sensitivities of nitro compounds obtained from drop weight impact tests. In practice, $\log h_{50\%}$ (with $h_{50\%}$ in cm) was considered since it has been successfully used to exhibit correlations in previous works [13,14,33,35,38].

The data set consisted in 60 nitramines, 50 nitroaliphatics and 51 nitroaromatics, i.e. 161 nitro compounds, presented in table 1. To allow the validation of models, it was divided into two sets. A training set was used to develop models and a validation set to estimate their predictive power. To achieve efficient validation, the validation set should be representative of the chemical diversity of molecules in the applicability domain of the model, i.e. included in the training set. Here, the partition between sets was done in each family of nitro compounds (nitramines, nitroaliphatics, nitroaromatics) to ensure the sufficient representativity of each family in the final model and to ensure an adequate number of molecules in each set for the development and validation of specific models. Moreover, sets were built to respect same distribution in experimental values in the training and validation sets. To this end, molecules were ordered by increasing experimental values in each

family and validation molecules were selected regularly (one molecule out of three). The remaining molecules constituted the training set.

Molecular descriptors

The molecular structures of the 161 nitro compounds of the data set were calculated using the density functional theory (DFT) in Gaussian03 package [40]. Geometries were optimized using the PBE0 functional [41] and the 6-31+G(d,p) basis set. This method has already been used in QSPR studies [42-45]. Moreover it was validated from geometrical and energetic considerations for studying the decomposition of nitro compounds [46,47]. Vibrational frequencies were computed at same calculation level to ensure that all stable species presented no imaginary frequency.

These molecular structures were then characterized by a series of descriptors. Four classes of descriptors can be considered. Constitutional descriptors exhibit the presence and number of specific features of molecules, e.g. numbers of atoms, bonds or chemical groups. Topological descriptors describe the 2D structure of the molecule based on its connectivity table. They characterize the size and shape of molecules, e.g. the Wiener index is well correlated with the boiling point of alkanes as it mathematically represents their degrees of branching [48]. Geometric descriptors stem for the 3D-structure (e.g. molecular volume). Quantum chemical descriptors include binding, energetic, electronic and thermodynamic information. They are particularly useful to characterize the reactivity of compounds.

Here, most of the descriptors resulted from the DFT calculated structures using CodessaPro software [49]. Detailed definitions and information about these descriptors can be found in [50]. Additional descriptors were also included since they have been identified in previous works to be related to the properties (e.g. their heats of decomposition [43,51]) of nitro energetic compounds or to characterize some particular features identified in the dataset. Finally, more than 300 descriptors were computed for each molecule in this study.

QSPR modeling

The QSPR approach relies on the principle that modifications of molecular structures undergo modifications of the properties of chemical compounds. It consists in developing mathematical relationships between experimental properties and descriptors of the molecular structures. To this end, various approaches can be used like genetic algorithms (GA) [52] or artificial neural networks (ANN) [53]. In this paper, multi linear regressions were computed. Such models follow the mathematic form in Eq. 1.

$$Y = a_0 + \sum_i a_i X_i \tag{1}$$

where Y is the property to predict, X_i are the molecular descriptors and a_i are the corresponding regression constants.

To achieve reliable and meaningful models, the statistical treatment has to select the set of n descriptors, among the more than 300 calculated ones, that leads to the development of the most predictive model. Here, the so-called "best multi linear regression" procedure (BMLR), as

implemented in CodessaPro software, was used. This method, already used successfully for the development of predictive models of nitro compounds [45] and described in ref [50], is based on the successive addition of descriptors in original two-parameter equations as long as the stability of models, evaluated based on the Fisher criterion F, increases. It finally provides the best model for each number of descriptors. The final model is chosen as the model presenting the best compromise between parameterization (i.e. the number of descriptors) and correlation.

Internal and external validations

To evaluate the performances of models, internal and external validations were performed. The goodness-of-fitting was first characterized by the coefficient of determination (R^2) and the root mean squared error (RMSE) between calculated and experimental values for the molecules of the training set. The pertinence of each descriptor in the model was checked based on a student t-test at a 95% level of confidence.

Then, the robustness of models was estimated by leave-one-out and leave-many-out cross validations (Q^2_{LOO}, Q^2_{10CV}, Q^2_{5CV}). Models were expected to present Q^2 values close to R^2 ones. Moreover, to ensure against chance correlations, Y-randomization tests [54,55] were computed. The values of properties were randomized 500 times and, at each iteration, new models were refitted. These new models are expected to present poor correlations compared to original ones. As proposed by Rücker et al.[54], chance correlation was considered at a 1% level of probability, by considering that the difference between the correlation of the original model and the average correlation of the random models issued from Y-randomisation (R^2_{YS}) should not exceed 2.3 times the standard deviation in R^2 for the new models (SD_{YS}).

Lastly, the predictive power of models was characterized by external validation on the correlation between predicted and experimental values for the molecules of the validation set. Performances of models were clarified based on the root mean square error ($RMSE_{EXT}$), the R^2_{EXT} and Q^2_{EXT} coefficients obtained from the predictions on the validation set. A summary of used statistical coefficients is proposed in table 2.

Applicability Domain (AD)

The domain of applicability of a model is the domain in which predictions are reliable. Indeed, by construction, a QSPR model is not applicable for all chemicals but only predictions for molecules that are similar to the ones of the training set can be considered as reliable.

In this paper, the domain of applicability of each model was defined by interpolation of the molecules of the training set, in the chemical space defined on the descriptors constituting the model. The Ambit Discovery software [56] was used to build these applicability domains to include 95% of the molecules of the training set based on their Euclidean distance to the mean in the descriptors space.

Hence, all external validation statistics have been computed again by excluding the molecules of the validation set that did not belong to the applicability domain of the models. These parameters were noted $RMSE_{IN}$, R^2_{IN}, Q^2_{IN}.

Moreover, for final application, it should be noted that models also have to be used in the range of experimental property values for the molecules in the training set. So, predictions calculated out of this range should be considered with particular attention.

Partial Least Squares

To assist discussion, partial least squares (PLS) analyses were also carried out using SIMCA P+ software [57]. This method consists in projecting an initial set of intercorrelated descriptors into a reduced set of orthogonal variables, called latent variables (LV). By construction, first projection represents the maximum of variance in the data set, taking into account the covariance of each descriptor with the impact sensitivity and additional projections aim to take into account the remaining variance. A LV was considered as necessary if it increased the robustness of the PLS analysis by 0.05. This approach allows the representation of the chemical space by plotting molecules on a 2D-graph based on the first two LVs (score plot) since these two LVs gathered most part of the variance in the dataset.

RESULTS

To predict the impact sensitivity of nitro compounds, two strategies were applied in this paper. In a first step, local models, dedicated to each family of nitro compounds (nitramines, nitroaliphatics and nitroaromatics), were developed. Prediction based on local models was successfully used by Papa et al. for the prediction of physico-chemical properties of polybrominated diphenyl ethers [8], Colombo et al. for the prediction of aqueous toxicity [58] and by Maunz et al. for chemical toxicity [59]. In particular, Colombo et al. proposed an approach based on the partition of the chemical space in subsets based on structural parameters (e.g. the number of aromatic rings, the maximum bond order for a C atom) and the application of a local model for each identified subset. More recently, Buchwald et al. [60] developed a more complex approach for the treatment of various large datasets (related to toxicological endpoints), in which the predictions are obtained by linear combination of the responses issued from different global and/or local models.

In a second step, a global model was built, to take into account of the whole diversity of compounds in the dataset (nitramines, nitroaliphatics and nitroaromatics) and the performances of this last model were compared to the predictions issued from local models.

Model for nitramines

The first developed model was dedicated to nitramine compounds. So, the BMLR procedure was applied to the 41 ones of the training set that range from 0.70 and 2.51 in $\log h_{50\%}$. Models with up to seven descriptors were computed and the best compromise between the correlation and the number of descriptors in the model was found for the four-parameter equation in Eq. 2.

$$\log h_{50\%} = 0.42 - 0.017\, OB + 0.06\, T_1^E + 50.1\, N_{O,avg} + 27.6\, N_{N,avg} \tag{2}$$

where OB is the oxygen balance as defined in the TDG regulation [61], T_1^E is the topological electronic index (for all pairs of atoms), $N_{O,avg}$ and $N_{N,avg}$ are the average nucleophilic reactivity indices for a O and a N atom, respectively.

The presence of the oxygen balance was expected since it has been already proposed to be correlated with the impact sensitivity [13,14,19,62]. In this equation, $N_{O,avg}$ and $N_{N,avg}$ are related to the reactivity of nitro groups that is critical in the decomposition process of nitro compounds [24]. The contribution of topological electronic index T_1^E in the model is less important. Nevertheless, it is not the most important descriptor, particularly regarding *OB* with a t-test of -15.5 vs. 5.0 for T_1^E.

The fit of this model (presented in Figure 1) is very good since the correlation between calculated and experimental values in the training set is high (R^2=0.919, RMSE=0.11). Moreover, the cross validation correlation coefficients are close to R^2 and stable when the partition size increases (Q^2_{LOO}=0.888, Q^2_{10CV}=0.878, Q^2_{5CV}=0.867), exhibiting good robustness. The model does not result from chance correlation, since the models generated from Y-randomization present low correlations (R^2_{YS}=0.103, SD_{YS}=0.071). Finally, predictions realized on the 19 nitramines of the validation set are in good agreement with experimental values (R^2_{EXT}=0.878, Q^2_{EXT}=0.853, $RMSE_{EXT}$=0.16), especially in the applicability domain (R^2_{IN}=0.845, Q^2_{IN}=0.817, $RMSE_{IN}$=0.16), as demonstrated in Table 3.

These performances are better than the ones of Wang's models for nitramines [38] in table 4. In particular, correlation is better than his MLR model (R^2=0.852). Moreover, the robustness and the predictive power of the new model are also significantly higher (Q^2_{LOO}=0.888 vs. 0.695, Q^2_{EXT}=0.853 vs. 0.670 and $RSME_{EXT}$=0.16 vs. 0.23). These better results, in particular in robustness, can be attributed to an over-parameterization of the models of Wang, based on 10 descriptors vs. only 5 in the present study for nearly the same number of molecules.

Model for nitroaliphatics

A second specific model was developed on the 34 nitroaliphatic compounds of the training set. A four-parameter model (in Eq. 3) was proposed to be the best compromise between correlation and number of descriptors [63].

$$\log h_{50\%} = -0.438 - 0.018\,OB + 4.07\,P_{Q\max-Q\min} + 28.5\,Q^2_{NO2,\max} + 4.79\,N_{O,\max} \tag{3}$$

where $P_{Qmax-Qmin}$ is the polarity parameter defined as the difference between the maximum and minimum charges in the molecule, $Q^2_{NO2,max}$ is the squared maximum charge for a nitro group (calculated from natural population analysis [64]) and $N_{O,max}$ is the maximum nucleophilic reactivity index for a O atom.

OB remains pertinent for the prediction of impact sensitivity of nitroaliphatic compounds. Moreover, the regression constants and t-test values for this descriptor are very similar in both models (0.017 vs. 0.018 and 15.5 vs. 14.8, for Eqs. 2 and 3 respectively). The three other descriptors are quantum chemical descriptors. In particular, $N_{O,max}$ 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. Besides, the electronic properties of nitro groups also strongly influence the last parameter, $P_{Qmax-Qmin}$.

The performances of the model are better than the MLR model of Wang for nitroaliphatic compounds [38]. The new equation is more correlated with R^2=0.929 (vs. 0.801 for Wang). Moreover, the models obtained after Y-randomization demonstrate low correlations (R^2_{YS}=0.123 and

SD$_{YS}$=0.076). So, this new model did not result from chance correlation. All property values in the training set, which define the AD of the model, ranged between 0.78 and 2.48 (log unit). Its predictive power, calculated for the 16 nitroaliphatic molecules of the validation set (see. Figure 2), remains high when considering the AD of the model with RSME$_{IN}$=0.19. Here again, robustness is good **(Q²$_{LOO}$=0.901)** when Wang's models clearly failed (Q²$_{LOO}$=0.512) due to possible over-parameterization.

Model for nitroaromatics

The last class of compounds in the data set concerned 51 nitroaromatics. A QSPR model was developed on the 34 nitroaromatic compounds of the training set. Among the 18 models issued from BMLR, the four-parameter equation in Eq. 4 demonstrated the best compromise between correlation and parameterization.

$$\log h_{50\%} = 17.4 + 0.69\, \varepsilon_{LUMO} - 8.4\, BO_{N,\max} - 1.7\, Q_{\min} - 25.4\, N_{C,\min} \tag{4}$$

where ε_{LUMO} is the energy of the lowest unoccupied molecular orbital (LUMO), $BO_{N,max}$ is the maximum bond order for a N atom, Q_{min} is the minimum partial charge and $N_{C,min}$ is the minimum nucleophilic reactivity index for a C atom.

The performances of this model are less important than the ones for nitramines and nitroaliphatics. Indeed, correlation and robustness are lower, with R²=0.812 and Q²$_{LOO}$=0.751, and, due to molecules out of the applicability domain, predictivity in terms of RMSE$_{EXT}$ exceeds the domain of values of logh$_{50\%}$ in the data set. Indeed, RMSE$_{EXT}$ of 3.502 was calculated whereas the values for the molecules of the data set are concentrated between 0.70 and 2.51. In particular, the predicted values for 2,4,6-trinitrobenzonitrile (aro-40) and 2,4,6-trinitrobenzaldoxime (aro-48) are negative and thus non physical (-9.67 and -0.59, respectively). These important deviations (clearly shown in Figure 3) made the statistical coefficients Q²$_{EXT}$ out of its classic range of values (-71.09). When considering only the applicability domain, this coefficient becomes meaningful, even if performances remain low in terms of prediction (R²$_{IN}$=0.541, Q²$_{IN}$=0.508, RMSE$_{IN}$=0.32). Besides, the models proposed by Wang [38] for nitroaromatics presented similar failure in prediction (R²$_{EXT}$=0.522, Q²$_{EXT}$=0.457, RMSE$_{EXT}$=0.27, for its MLR model in Table 4).

The complexity of the mechanisms involved in the decomposition of nitroaromatic compounds offers an explanation to such difficulty. Indeed, different reaction paths have been identified for the decomposition of nitroaromatic compounds upon their molecular structures [65]. For instance, particular primary paths were exhibited for nitroaromatics presenting a methyl group in ortho position to a nitro group [47] compared to nitroaromatics without any ortho groups [46]. Such consideration was considered by Kamlet [14] who proposed two models as a function of the presence of CH containing groups in ortho position to a nitro leading to good correlations (r=0.97 and 0.96 for the compounds without and with α-CH linkage, respectively). Besides, similar subdivision was proposed in previous work for the prediction of heats of decomposition [45]. These two subdivisions were also considered in the present study. However, no reliable model was obtained when considering only non-ortho-substituted or only ortho-substituted nitroaromatic compounds. The development of a model dedicated to the nitro compounds presenting α-CH linkage was not

possible since the number of available data (only 17 compounds) was not sufficient for partition between training and validation sets.

Projection methods, e.g. PLS, are efficient tools to identify specific classes of molecules in databases, and were used here to identify possible particular clusters of compounds among nitroaromatics. Indeed, the presence of specific groups of molecules in data sets, generally due to particular molecular features, could explain the difficulty to achieve reliable models. So, a PLS analysis was performed on the entire set of 51 nitroaromatics based on the entire set of descriptors, after exclusion of all descriptors presenting missing values or no variance, i.e. 313 descriptors. The score plot, in Figure 4, obtained by projection of the molecules on the first two LVs, proposes an illustration of the distribution of molecules as a function of the descriptors and their covariance with the impact sensitivity. No distinction was observed between ortho and non-ortho substituted compounds as in previous works [45], but clusters were identified as a function of other molecular features: number of aromatic cycles and presence of a $C(NO_2)_3$ fragment. In particular, this last feature could be related to a particular decomposition path during the decomposition process.

To go further in the development of reliable models, a first potential strategy would consist in developing different models for each identified clusters in this PLS analysis. Unfortunately, the available data set was not large enough. Indeed, the largest cluster contained only 27 molecules, which is not sufficient to be divided into sufficiently large training and validation sets.

In another approach, additional experimental measurements may lead to a dataset producing a better distribution of the molecules in the chemical space, by performing. In this paper, the data set was extracted from a single reference taken from literature. But, impact sensitivity is very dependent on experimental apparatus and other data from other sources may not be consistent. Nevertheless, in this paper, the data reported in reference [39] made it possible to extend the data set to nitramines and nitroaliphatics, as proposed in the following global analysis.

Global model

In the perspective of a more global approach towards the prediction of impact sensitivities of nitro compounds, the complete data set was considered. It contained 161 molecules and was balanced between the three families of nitro compounds: 61 nitramines, 51 nitroaromatics and 50 nitroaliphatics. Thus, this data set could be expected to be adapted to the development of a global model.

Moreover, it could allow further predictions for nitroaromatic compounds by introducing additional compounds that could better represent the diversity of mechanisms that could be involved in the decomposition of nitro compounds as explained in previous section.

This global model was developed on the entire training set, i.e. 108 molecules. Using the BMLR method, regressions with up to 15 descriptors were computed and a five-parameter regression (in Eq. 5) was selected.

$$\log h_{50\%} = 0.60 - 0.013\,OB + 42{,}9\,N_{O,avg} - 2.4\,Q_{min} + 0.011\,WPSA3 + 0.15\,n_{NH2} \qquad (5)$$

9

where OB is the oxygen balance as defined in the TDG regulation [61], $N_{O,avg}$ is the average nucleophilic reactivity index for a O atom, Q_{min} is the minimum partial charge, WPSA3 is the weighted partial positive charged surface area (order 3) obtained via calculated Mulliken charges [66] and n_{NH2} is the number of amino groups.

Some chemical interpretation can be highlighted. As in the models for nitramines and nitroaliphatics, the significance of OB (t-test=-16.2) is consistent with the correlations made by Kamlet [13,14] and Lothrop [62]. Moreover, $N_{O,avg}$ characterizes the reactivity of the O atoms in nitro groups that are central in the decomposition process of nitro compounds. Moreover, if the pertinence of the descriptors WPSA3 and Q_{min}, are less easy to clarify, they are based on the distribution of charge in molecules, which is strongly influenced by the presence of nitro groups. Lastly, the presence of the number of amino groups did not appear so likely interpretable. Nevertheless, it is the less significant descriptor in the equation (t-test=2.6). The development of a model excluding n_{NH2} was tested but the performance of the models decreased, in particular for the amino molecules of the validation set.

The performances of this model are satisfactory (as presented in figure 5). Indeed, it is not only well correlated ($R^2=0.816$) and robust ($Q^2_{LOO}=0.793$), but also predictive ($R^2_{EXT}=0.759$), in particular in its applicability domain ($R^2_{IN}=0.750$). In terms of predicted impact sensitivity, predictions between 0.70 and 2.51 (log unit) can be considered as reliable. This model is more predictive than the global MLR model of Wang [38] ($R^2_{EXT}=0.740$). It is less reliable than the models developed previously in this paper for nitramines and nitroaliphatics but better than the one for nitroaromatic compounds. The same observations can be made for the models of Wang.

To complete the comparison between local and global models, the difference between the values obtained from Eqs. 2-4 and experimental results was computed for the molecules of the validation set (taking into account ADs) and a better predictive power was observed for a consensus modelling based on local models ($R^2_{IN}=0.829$) than for the global model ($R^2_{IN}=0.750$), as shown in table 5. Moreover, it can be noticed that if the predictive power of the global model is significant, it failed for nitroaromatic compounds with only $R^2_{IN}=0.227$, contrary to the hope of improvement that could be issued from the extension of the dataset in terms of chemical diversity. Local models lead to better predictivity than the global one for nitroaliphatics ($R^2_{IN}=0.877$ vs 0.816) and similar results are obtained for nitramines ($R^2_{IN}=0.845$ vs 0.835). So, these two approaches lead to the same conclusions. In both cases, a failure was observed for nitroaromatic compounds whereas good predictions were obtained for nitramines and nitroaliphatics (with higher reliability for local models than for the global one, as it could have been expected).

Finally, theses two approaches are complementary. Indeed, local models give the best predictions for some molecules, e.g. for the 1,1,1,3-Tetranitrobutane (ali-35) with an error of 0.12 compared with an error of 0.31 for the global model, whereas the global model is more predictive for others, such as trinitroethyl-5,5-dinitro-3-nitrazahexanoate (ami-49) and 2,2,2-trinitroethyl-4,4,4-trinitrobutyrate (ali-40) with deviations of 0.01 and 0.03 vs. 0.19 and 0.14 for the local models, respectively. So, global and local approaches are complementary and they could be used in a consensus by considering the prediction from both models to obtain a more robust estimation of the sensitivity of studied compounds.

CONCLUSION

In this study, four QSPR models were developed and validated in view of predicting the impact sensitivities of nitro compounds, based on multilinear regressions and quantum chemical calculations. The data set used was well balanced between three classes of nitro compounds: nitramines, nitroaliphatics and nitroaromatics. First, three local models were developed for each of the three families. Then, a fourth model was developed on a global approach using all the nitro compounds of the data set.

All models were constructed according to all OECD principles for the validation of QSAR/QSPR models for regulatory use. Indeed, the endpoint was clearly defined as the impact sensitivity from drop weight impact test. Algorithms were completely defined from the model equations to the calculation of initial molecular structures. Moreover, the applicability domain was determined based on the chemical space in the training set and performances were characterized by internal and external validation procedures to measure the goodness-of-fit, the robustness and the predictive power of models. Furthermore, descriptors in models were pertinent from a chemical point of view since most of them are related (directly or indirectly) to the properties of nitro groups that are critical in the reaction mechanisms involved in the decomposition of nitro compounds.

For both nitramines and nitroaliphatics, local models demonstrated high performances, in particular in terms of predictivity ($RMSE_{IN}$=0.16 and 0.19 respectively), whereas analysis failed for nitroaromatics due to the too large diversity of structures regarding the size of the data set. Concerning the global model, it demonstrated significant predictive power with $RMSE_{IN}$=0.21 in its applicability domain. Regarding the predictions of impact sensitivities for nitroaromatics, both local and global models failed. Even though the performance of this last model was good, local models could be preferred since their performances were higher in the case of nitramines and nitroaliphatics. A consensus approach based on both local and global approaches could also be pertinent to improve confidence in the predictions.

REFERENCES

1.   Globally harmonized system of classification and labelling of chemicals (GHS), United Nations, Geneva - New York, 2009.
2.   Recommendations on the Transport of Dangerous Goods: Manual of Tests and Criteria, United Nations, Geneva/New York, 2011.
3.   E. Papa, S. Kovarich, and P. Gramatica, Development, Validation and Inspection of the Applicability Domain of QSPR Models for Physicochemical Properties of Polybrominated Diphenyl Ethers, QSAR Comb. Sci. 28 (2009), 790-796.
4.   M. Cocchi, P.G. De Benedetti, R. Seeber, L. Tassi, and A. Ulrici, Development of Quantitative Structure-Property Relationships Using Calculated Descriptors for the Prediction of the Physicochemical Properties (n_D, ρ, bp, ε, η) of a Series of Organic Solvents, J. Chem. Inf. Comput. Sci. 39 (1999), 1190-1203.
5.   A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, and D.A. Dobchev, Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction, Chem. Rev. 110 (2010), 5714-5789.
6.   F. Gharagheizi, A QSPR model for estimation of lower flammability limit temperature of pure compounds based on molecular structure, J. Hazard. Mat. 169 (2009), 217-220.
7.   G. Fayet, P. Rotureau, L. Joubert, and C. Adamo, Predicting explosibility properties of chemicals from Quantitative Structure-Property Relationships, Process Saf. Prog. 29 (2010), 359-371.

8.   E. Papa, S. Kovarich, and 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.

9.   E. Papa, and 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.

10.   B. Bhhatarai, W. Teetz, T. Liu, T. Öberg, N. Jeliazkova, N. Kochev, O. Pukalov, I.V. Tetko, S. Kovarich, E. Papa, and P. Gramatica, CADASTER QSPR Models for Predictions of Melting and Boiling Points of Perfluorinated Chemicals, Mol. Inf. 30 (2011), 189-204.

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).

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.J. Kamlet, The Relationship of Impact Sensitivity with Structure of Organic High Explosives. I. Polynitroaliphatic Explosives, Sixth Symposium (International) on Detonation, Coronado, CA, (1976), pp. 69-72.

14.   M.J. Kamlet, and H.G. Adolph, The relationship of Impact Sensitivity with Structure of Organic High Explosives. II. Polynitroaromatic explosives, Propel. Explos. 4 (1979), 30-34.

15.   B.M. Rice, and 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.

16.   J. Mullay, A Relationship between Impact Sensitivity and Molecular Electronegativity, Propel. Explos. Pyrotech. 12 (1987), 60-63.

17.   J. Mullay, A Relationship between Impact Sensitivity and Molecular Electronic Structure, Propel. Explos. Pyrotech. 12 (1987), 121-124.

18.   C. Zhang, Y. Shu, Y. Huang, X. Zhao, and H. Dong, Investigation of Correlation between Impact Sensitivities and Nitro Group Charges in Nitro Compounds, J. Phys. Chem. B 109 (2005), 8978-8982.

19.   C. Cao, and S. Gao, Two Dominant Factors Influencing the Impact Sensitivities of Nitrobenzenes and Saturated Nitro Compounds, J. Phys. Chem. B 111 (2007), 12399-12402.

20.   P. Politzer, and J.S. Murray, Relationships between dissociation energies and electrostatic potentials of C-NO2 bonds: applications to impact sensitivities, J. Molec. Struct. 376 (1996), 419-424.

21.   J.S. Murray, P. Lane, P. Politzer, and P.R. Bolduc, A relationship between impact sensitivity and the electrostatic potentials at the midpoints of C-NO2 bonds in nitroaromatics, Chem. Phys. Lett. 168 (1990), 135-139.

22.   F.J. Owens, Calculation of energy barriers for bond rupture in some energetic molecules, J. Mol. Struct. (THEOCHEM) 370 (1996), 11-16.

23.   X.-S. Song, X.-L. Cheng, X.-D. Yang, and B. He, Relationship between the Bond Dissociation Energies and Impact Sensitivities of Some Nitro-Explosives, Propel. Explos. Pyrotech. 31 (2006), 306-310.

24.   V.L. Korolev, T.S. Pivina, A.A. Porollo, T.V. Petukhova, A.B. Sheremetev, and V.P. Ivshin, Differentiation of the molecular structure of nitro compounds as the basis for simulation of their thermal destruction processes, Russ. Chem. Rev. 78 (2009), 945-969.

25.   H. Nefati, J.-M. Cense, and J.-J. Legendre, Prediction of the Impact Sensitivity by Neural Networks, J. Chem. Inf. Comput. Sci. 36 (1996), 804-810.

26.   S.G. Cho, K.T. No, E.M. Goh, J.K. Kim, J.H. Shin, Y.D. Joo, and S. Seong, Optimization of Neural Networks Architecture for Impact Sensitivity of Energetic Molecules, Bull. Korean Chem. Soc. 26 (2005), 399-408.

27. Z. Jun, C. Xin-lu, H. Bi, and Y. Xiang-dong, Neural networks study on the correlation between impact sensitivity and molecular structures for nitramine explosives, Struct. Chem. 17 (2006), 501-507.

28. J. Li, A multivariate relationship for the impact sensitivities of energetic N-nitrocompounds based on bond dissociation energy, J. Hazard. Mater. 174 (2010), 728-733.

29. M.H. Keshavarz, Prediction of impact sensitivity of nitroaliphatic, nitroaliphatic containing other functional groups and nitrate explosives, J. Hazard. Mater. 148 (2007), 648-652.

30. M.H. Keshavarz, H.R. Pouretedal, and A. Semnani, Novel correlation for predicting impact sensitivity of nitroheterocyclic energetic molecules, J. Hazard. Mater. 141 (2007), 803-807.

31. M.H. Keshavarz, and M. Jaafari, Investigation of the Various Structure Parameters for Predicting Impact Sensitivity of Energetic Molecules via Artificial Neural Network, Propel. Explos. Pyrotech., 31 (2006), 216-225.

32. M.H. Keshavarz, and H.R. Pouretedal, Simple empirical method for prediction of impact sensitivity of selected class of explosives, J. Hazard. Mater. 124 (2005), 27-33.

33. M.H. Keshavarz, A. Zali, and A. Shokrolahi, A simple approach for predicting impact sensitivity of polynitroheteroarenes, J. Hazard. Mater. 166 (2009), 1115-1119.

34. G.T. Afanas'ev, T.S. Pivina, and D.V. Sukhachev, Comparative characteristics of some experimental and computational methods for estimating Impact Sensitivity Parameters of Explosives, Propel. Explos. Pyrotech., 18 (1993), 309-316.

35. N.R. Badders, C. Wei, A.A. Aldeeb, W.J. Rogers, and M.S. Mannan, Predicting the Impact Sensitivities of Polynitro Compounds Using Quantum Chemical Descriptors, J. Energ. Mater. 24 (2006), 17 - 33.

36. J.A. Morrill, and E.F.C. Byrd, Development of quantitative structure–property relationships for predictive modeling and design of energetic materials, J. Mol. Graph. Model. 27 (2008), 349-355.

37. M.H. Keshavarz, Simple Relationship for Predicting Impact Sensitivity of Nitroaromatics, Nitramines, and Nitroaliphatics, Propel. Explos. Pyrotech., 35 (2010), 175-181.

38. R. Wang, J. Jiang, Y. Pan, H. Cao, and 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.

39. C.P. Storm, and J.R. Stine, *Sensitivity Relationships in Energetic Materials*, in S.N. Bulusu (Ed.), Chemistry and Physics of Energetic Materials. Kluwer Academic Publishers, Netherlands, 1990, p. 605-630.

40. 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, and J.A. Pople, Gaussian03. Gaussian Inc., Wallington CT, 2004.

41. C. Adamo, and V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, J. Chem. Phys. 110 (1999), 6158-6170.

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

43. G. Fayet, L. Joubert, P. Rotureau, and 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.

44. G. Fayet, P. Rotureau, L. Joubert, and C. Adamo, QSPR Modeling of Thermal Stability of Nitroaromatic Compounds: DFT vs. AM1 Calculated Descriptors, J. Mol. Model. 16 (2010), 805-812.

45. G. Fayet, P. Rotureau, L. Joubert, and 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.

46. G. Fayet, L. Joubert, P. Rotureau, and C. Adamo, Theoretical Study of the Decomposition Reactions in Substituted Nitrobenzenes, J. Phys. Chem. A 112 (2008), 4054-4059.

47. G. Fayet, L. Joubert, P. Rotureau, and C. Adamo, A theoretical study of the decomposition mechanisms on substituted ortho-nitrotoluenes, J. Phys. Chem. A 113 (2009), 13621-13627.

48. H. Wiener, Structural Determination of Paraffin Boiling Points, J. Am. Chem. Soc. 69 (1947), 17-20.

49. CodessaPro. University of Florida, 2002.

50. M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, 2000.

51. G. Fayet, P. Rotureau, L. Joubert, and C. Adamo, On the prediction of thermal stability of nitroaromatic compounds using quantum chemical calculations, J. Hazard. Mater. 171 (2009), 845-850.

52. R. Leardi, Genetic algorithms in chemometrics and chemistry: a review, J. Chemometr. 15 (2001), 559-569.

53. J. Gasteiger, and J. Zupan, Neural Networks in Chemistry, Angew. Chem. Int. Ed. Engl. 32 (1993), 503-527.

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

55. F. Lindgren, B. Hansen, W. Karcher, M. Sjöström, and L. Eriksson, Model validation by permutation tests: Applications to variable selection, J. Chemometrics 10 (1996), 521-532.

56. J. Jaworska, and N. Jeliazkova, Ambit Discovery 2007.

57. SIMCA P+, Version 12.0. Umetrics AB, Umeå, Sweden, 2009.

58. A. Colombo, E. Benfenati, M. Karelson, and U. Maran, The proposal of architecture for chemical splitting to optimize QSAR models for aquatic toxicity, Chemosphere 72 (2008), 772-780.

59. A. Maunz, and C. Helma, Prediction of chemical toxicity with local support vector regression and activity-specific kernels, SAR QSAR Environ. Res. 19 (2008), 413-431.

60. F. Buchwald, T. Girschick, M. Seeland, and S. Framer, Using local models to improve (Q)SAR predictivity, Mol. Inform. 30 (2011), 205-218.

61. Recommendations on the Transport of Dangerous Goods: Manual of Tests and Criteria, United Nations, Geneva/New York, 2009.

62. W.C. Lothrop, and G.R. Handrick, The Relationship between Performance and Constitution of Pure Organic Explosive Compounds, Chem. Rev. 44 (1949), 419-445.

63. V. Prana, G. Fayet, P. Rotureau, and C. Adamo, Predictive QSPR models for impact sensitivity of nitroaliphatic compounds, manuscript in preparation.

64. A.E. Reed, and F. Weinhold, Natural localized molecular orbitals, J. Chem. Phys. 83 (1985), 1736-1740.

65. T.B. Brill, and K.J. James, Kinetics and mechanisms of thermal decomposition of nitroaromatic explosives, Chem. Rev. 93 (1993), 2667-2692.

66. R.S. Mulliken, Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I, J. Chem. Phys. 23 (1955), 1833-1840.

Table 1 – Experimental and predicted impact sensitivities (logh$_{50\%}$) from local and global models (Eqs 2-5)

ID	compounds	Experimental	Predicted			
		Ref. [39]	Nitramines Eq. 2	Nitroaliphatics Eq. 3	Nitroaromatics Eq. 4	Global Eq. 5
Training set						
ami-01	N,N'-Dinitro-methanediamine	1.11	1.21	-	-	1.21
ami-02	N-Nitro-N-methyl-formamide	2.51	2.52	-	-	2.44
ami-03	Methyl-2,2,2-trinitro-ethylnitramine	0.95	1.00	-	-	0.94
ami-04	Trinitroethyl-nitroguanidine	1.18	1.00	-	-	0.99
ami-05	Cyclotrimethylene-trinitramine	1.41	1.35	-	-	1.34
ami-06	N-Methyl-N,N'-dinitro-1,2-ethanediamine	2.06	2.08	-	-	1.89
ami-07	Trinitroethyl-cyanomethylnitramine	1.04	1.09	-	-	1.09
ami-08	Bis-(2,2,2-trinitroethyl)-nitramine	0.70	0.56	-	-	0.60
ami-09	N,N'-Dimethyl-N,N'-dinitrooxamide	1.90	1.86	-	-	1.93
ami-10	N-Nitro-N-(trinitroethyl)-glycinamide	1.23	1.17	-	-	1.53
ami-11	Cyclotetramethylene-tetranitramine	1.46	1.38	-	-	1.28
ami-12	N,N'-Dinitro-N-[2-(nitramino)ethyl]-1,2-ethanediamine	1.59	1.71	-	-	1.73
ami-13	1,3,3,5,5-Pentanitropiperidine	1.15	1.16	-	-	1.07
ami-14	2,2,2-Trinitroethyl-3',3',3'-trinitropropyl-nitramine	0.78	0.80	-	-	0.79
ami-15	Trinitroethyl-N-ethyl-N-nitro-carbamate	1.28	1.44	-	-	1.54
ami-16	Trinitroethyl-2-methoxy-ethylnitramine	1.62	1.44	-	-	1.88
ami-17	N,N'-3,3-Tetranitro-1,5-pentanediamine	1.54	1.59	-	-	1.60
ami-18	2,2,2-Trinitroethyl-N-(2,2,2-trinitroethyl)-nitramino acetate	0.95	0.91	-	-	1.05
ami-19	2,2,2-Trinitroethyl-4-nitrazavalerate	1.54	1.46	-	-	1.55
ami-20	Trinitropropyle-(2,2-dinitropropyl)-nitramine	1.23	1.20	-	-	1.17
ami-21	2',2',2'-Trinitroethyl-2,5-dinitrazahexanoate	1.18	1.25	-	-	1.39
ami-22	2,2,2-Trinitroethyl-3,3-dinitrobutyl nitramine	1.30	1.20	-	-	1.17
ami-23	N,N'-Dinitro-N,N'-bis[2-(nitramino)ethyl]-1,2-ethanediamine	1.72	1.81	-	-	1.87

ID	Name					
ami-24	2,2-Dinitropropyl-5,5,5-trinitro-2-nitrazapentanoate	1.20	1.29	-	-	1.45
ami-25	2,2,2-Trinitroethyl-2,5,5-trinitro-2-azahexanoate	1.34	1.31	-	-	1.39
ami-26	2,2,2-Trinitroethyl-2,4,6,6-tetranitro-2,4-diazaheptanoate	1.26	1.16	-	-	1.28
ami-27	N,N'-Dinitro-N,N'-bis-(3,3,3-trinitropropyl)-oxamide	0.95	1.06	-	-	1.09
ami-28	2,2,6,9,9-Pentanitro-4-oxa-5-oxo-6-azadecane	1.67	1.67	-	-	1.86
ami-29	N-(2,2-Dinitrobutyl)-N-2,2-trinitro-1-butanamine	1.90	1.92	-	-	1.86
ami-30	N,N'-Dinitromethylene-bis-(4,4,4-trinitro)-butyramide	1.11	1.22	-	-	1.31
ami-31	1,1,1,5,7,10,14,14,14-Nonanitro-3,12-dioxa-4,11-dioxo-5,7,10-triazatetradecane	1.04	1.11	-	-	1.13
ami-32	1,1,1,4,6,6,9,11,11,11-Decanitro-4,8-diazaundecane	1.04	1.12	-	-	1.02
ami-33	1,1,1,3,6,6,9,11,11,11-Decanitro-3,9-diazaundecane	1.00	1.11	-	-	0.99
ami-34	Bis-(2,2,2-trinitroethyl)-3,6-dinitraza-1,7-heptanedioate	1.46	1.38	-	-	1.68
ami-35	Bis-(2,2,2-trinitroethyl)-3,6-dinitraza-1,8-octanedioate	1.46	1.39	-	-	1.44
ami-36	Bis-(trinitroethyl)-5,5-dinitro-2,8-dinitrazanonadioate	1.08	1.26	-	-	1.31
ami-37	1,4-Bis-(5,5,5-trinitro-2-nitrazapentanoate)-2-butyne	1.20	1.32	-	-	1.58
ami-38	1,1,1,18,18,18-Hexanitro-3,16-dioxa-4,15-dioxo-5,8,11,14-tetranitrazaoctadecane	1.28	1.37	-	-	1.35
ami-39	1,1,1,3,6,6,8,10,10,13,15,15,15-Tridecanitro-3,8,13-triazapentanedecane	1.36	1.26	-	-	1.05
ami-40	2,2-Dinitropropanediol-bis-(5,5-dinitro-2-nitraza-hexanoate)	2.14	1.79	-	-	1.86
ali-01	1,1,1,3,5,5,5-Heptanitropentane	0.90	-	0.84	-	0.89
ali-02	1,1,1,6,6,6-Hexanitro-3-hexyne	0.85	-	0.90	-	0.98
ali-03	3,3,4,4-Tetranitrohexane	1.90	-	1.89	-	1.96
ali-04	2,2,4,4,6,6-Hexanitroheptane	1.46	-	1.41	-	1.45
ali-05	2,2,4,6,6-Pentanitroheptane	1.75	-	1.91	-	1.88
ali-06	2,2,2-Trinitroethyl-carbamate	1.26	-	1.30	-	1.36
ali-07	2,2-Dinitro-1,3-propane-diol	2.04	-	2.16	-	1.95
ali-08	Methyl-2,2,2,-trinitro-ethyl carbonate	1.45	-	1.56	-	1.54
ali-09	4,4,4-Trinitrobutyramide	1.60	-	1.80	-	1.90

ali-10	Bis-(2,2,2 trinitroethyl)-carbonate	1.20	-	1.02	-	0.99
ali-11	Bis-(trinitroethoxy)-methane	1.23	-	1.24	-	1.19
ali-12	N,N'-Bis-(2,2,2-trinitroethyl)-urea	1.23	-	1.19	-	1.06
ali-13	Ethyl-2,2,2-trinitro-ethylcarbonate	1.91	-	1.88	-	1.84
ali-14	Bis-(trinitroethyl)-oxalate	1.18	-	1.03	-	1.14
ali-15	Bis-(trinitroethyl)-oxamide	1.11	-	1.18	-	1.15
ali-16	N-Trinitroehyl-4,4,4-trinitrobutyramide	1.26	-	1.21	-	1.27
ali-17	1,5-Bis-(trinitroethyl)-biuret	1.38	-	1.19	-	1.21
ali-18	N-(t-Butyl)-trinitro-acetamide	2.04	-	2.13	-	2.04
ali-19	1,1,1,7,7,7-Hexanitroheptanone-4	1.53	-	1.51	-	1.45
ali-20	2,2-Dinitropropyl-trinitrobutyrate	2.18	-	2.01	-	1.75
ali-21	2,2,2-Trinitroethyl-4,4-dinitrovalerate	1.85	-	1.88	-	1.72
ali-22	Bis-(2,2-dinitropropyl)-carbonate	2.48	-	2.33	-	1.93
ali-23	Bis-(trinitropropyl)-urea	1.36	-	1.24	-	1.44
ali-24	Bis-(trinitroethyl)fumarate	1.15	-	1.38	-	1.36
ali-25	Trinitroethyl-bis-(trinitroethoxy)-acetate	0.78	-	1.03	-	1.07
ali-26	4,4,4-Trinitrobutyric anhydride	1.48	-	1.55	-	1.49
ali-27	2,2,2-Trinitroerhyl-4,4-dinitrohexanoate	2.14	-	2.07	-	1.95
ali-28	Nitroisobutyl-4,4,4-trinitrobutyrate	2.45	-	2.42	-	2.36
ali-29	Tetrabis-(2,2,2-trinitro-ethyl)-orthocarbonate	0.85	-	0.86	-	0.83
ali-30	Methylene-bis-(4,4,4-trinitrobutyramide)	2.05	-	1.92	-	1.78
ali-31	Ethylene-bis-(4,4,4-trinitrobutyrate)	2.08	-	1.90	-	1.84
ali-32	N,N'-Bis-(2,2-binitropropyl)-4,4,4-trinitro-butyramide	1.86	-	1.83	-	1.74
ali-33	2,2-Dinitropropane-1,3-1,3-diol-bis-(4,4,4-trinitrobutyrate)	1.70	-	1.71	-	1.57
ali-34	Bis-(2,2,2-trinitroethyl)-4,4,6,6,8,8-hexanitro-undecanedioate	1.51	-	1.70	-	1.56
aro-01	Hexanitrobenzene	1.08	-	-	0.96	0.90
aro-02	Benzotrifuroxan	1.70	-	-	1.72	1.60
aro-03	1,3,5-Trinitrobenzene	2.00	-	-	1.81	1.99

aro-04	Picric acid	1.94	-	-	1.75	1.78
aro-05	2,4,6-Trinitroresorcinol	1.63	-	-	1.49	1.69
aro-06	2,4,6-Trinitrophloroglucinol	1.43	-	-	1.67	1.52
aro-07	2,3,4,6-Tetranitroaniline	1.61	-	-	1.63	1.52
aro-08	2,4-Dinitroresorcinol	2.47	-	-	2.25	2.11
aro-09	2,4,6-Trinitro-3-aminophenol	2.14	-	-	2.04	1.86
aro-10	1-Hydroxy-3,5-diamino-2,4,6-trinitrobenzene	2.08	-	-	2.09	2.06
aro-11	2,4,6-Trinitrobenzoic acid	2.04	-	-	2.07	1.92
aro-12	2,4,6-Trinitroanisole	2.28	-	-	2.36	2.09
aro-13	1,3-Dimethoxy-2,4,6-trinitrobenzene	2.40	-	-	2.44	2.25
aro-14	2',2',2'-Trinitroethyl-2,4,6-trinitrobenzoate	1.38	-	-	1.44	1.42
aro-15	2',2',2'-Trinitroethyl-3,5-dinitrobenzoate	1.86	-	-	1.61	1.57
aro-16	2',2',2'-Trinitroethyl-3,5-dinitrosalicylate	1.65	-	-	1.71	1.54
aro-17	3-Hydroxy-2,2',4,4',6,6'-hexanitrobiphenyl	1.62	-	-	1.73	1.73
aro-18	3,3'-Dihydroxy-2,2',4,4',6,6'-hexanitrobiphenyl	1.60	-	-	1.68	1.65
aro-19	2,2',4,4',6-Pentanitrobenzophenone	1.73	-	-	1.64	1.58
aro-20	3,3'-Diamino-2,2',4,4',6,6'-hexanitrobiphenyl	2.12	-	-	1.95	1.97
aro-21	2,2',2'',4,4'',6,6',6''-Octanitro-p-terphenyl	1.77	-	-	1.79	1.71
aro-22	2,4,6-Trinitrobenzaldehyde	1.56	-	-	1.89	1.90
aro-23	2,4,6-Trinitrotoluene	2.20	-	-	2.22	2.06
aro-24	1-Dinitromethyl-3-nitrobenzene	2.02	-	-	2.01	1.92
aro-25	2,4,6-Trinitrobenzyl alcohol	1.72	-	-	1.78	2.19
aro-26	2,4,6-Trinitro-m-cresol	2.28	-	-	2.03	2.09
aro-27	1-(2,2,2-Trinitroethyl)-2,4,6-trinitrobenzene	1.11	-	-	1.25	1.25
aro-28	2,4,6-Trinitrostyrene	1.51	-	-	1.56	2.03
aro-29	1-(2,2,2-Trinitroethyl)-2,4-dinitrobenzene	1.49	-	-	1.47	1.67
aro-30	3,5-Dimethyl-2,4,6-trinitrophenol	1.89	-	-	2.18	2.35
aro-31	1-(3,3,3-Trinitropropyl)-2,4,6-trinitrobenzene	1.32	-	-	1.47	1.42
aro-32	3-Methyl-2,2',4,4',6-pentanitrobiphenyl	2.16	-	-	2.12	2.17

aro-33	Hexanitrostilbene	1.59	-	-	1.83	1.86
aro-34	3,3'-Dimethyl-2,2',4,4',6,6'-hexanitrobiphenyl	2.13	-	-	1.91	1.92
Validation set						
ami-41	N,N'-Dinitro-1,2-ethanediamine	1.53	1.60	-	-	1.52
ami-42	N-Methyl-N-nitro-(trinitroethyl)-carbamate	1.23	1.03	-	-	1.30
ami-43	N,N'-(Bis-2,2,2-trinitroethyl)-N,N'-dinitromethanediamine	0.70	0.73	-	-	0.68
ami-44	N-Methyl-N'-trinitroethyl-N,N'-dinitro-1,2-ethanediamine	1.04	1.26	-	-	1.26
ami-45	N-Nitro-N-(3,3,3-trinitropropyl)-2,2,2-trinitroethyl carbamate	0.95	0.88	-	-	1.05
ami-46	N-(2,2-Dinitropropyl)-N-2,2-trinitro-1-propanamine	1.46	1.40	-	-	1.37
ami-47	1,7-Dimethoxy-2,4,6-trinitrazaheptane	2.22	2.01	-	-	2.39
ami-48	Bis-(trinitroethyl)-2,4-dinitrazapentanedioate	1.00	0.90	-	-	0.98
ami-49	Trinitroethyl-5,5-dinitro-3-nitrazahexanoate	1.40	1.21	-	-	1.41
ami-50	N-Nitro-N,N'-bis-(trinitropropyl)-urea	1.32	1.10	-	-	1.16
ami-51	Bis-(2,2,2-trinitroethyl)-3-nitrazaglutarate	1.15	1.18	-	-	1.30
ami-52	Bis-(trinitroethyl)-2,4,6-trinitrazaheptanedioate	1.11	0.94	-	-	1.05
ami-53	1,1,1,3,6,9,11,11,11-Nonanitro-3,6,9-triazaundecane	1.08	1.10	-	-	1.06
ami-54	N,N'-Dinitro-N,N'-bis-(3-nitrazabutyl)-oxamide	1.95	1.86	-	-	2.01
ami-55	2,2,4,7,9,9-Hexanitro-4,7-diazadecane	1.86	1.63	-	-	1.67
ami-56	Bis-(5,5,5-trinitro-3-nitrazapentanoyl)-methylene dinitramine	1.18	1.01	-	-	0.94
ami-57	Bis-(trinitroethyl)-2,5,8-trinitrazanonanedioate	1.23	1.24	-	-	1.29
ami-58	1,1,1,3,6,9,12,14,14,14-Decanitro-3,6,9,12-tetrazatetradecane	1.28	1.30	-	-	1.29
ami-59	2,2,4,7,7,10,12,12-Octanitro-4,10-diazatridecane	1.64	1.68	-	-	1.67
ami-60	2,2,4,7,7,9,12,12-Octanitro-5,9-diazatridecane	1.57	1.71	-	-	1.60
ali-35	1,1,1,3-Tetranitrobutane	1.52	-	1.40	-	1.83
ali-36	1,1,1,6,6,6-Hexanitro-3-hexene	1.23	-	1.04	-	1.03
ali-37	Methylene-bis-N,N'-(2,2,2-trinitroacetamide)	0.95	-	0.73	-	0.88

ali-38	5,5,5-Trinitropentanone-2	2.10	-	2.21	-	2.00
ali-39	N-(2-Propyl)-trinitroacetamide	2.05	-	1.83	-	1.80
ali-40	2,2,2-Trinitroethyl-4,4,4-trinitrobutyrate	1.26	-	1.43	-	1.29
ali-41	Trinitroethyl-2,2-dinitropropylcarbonate	1.18	-	1.49	-	1.40
ali-42	Tris-(2,2,2-trinitroethyl)-orthoformate	0.85	-	0.88	-	0.97
ali-43	Methylene-bis-(trinitroethyl)-carbamate	1.43	-	1.25	-	1.33
ali-44	2,2-Dinitropropyl–4,4,4-trinitrobutyramide	1.86	-	1.85	-	1.66
ali-45	Bis-(1,1,1-trinitro-2-propyl)-urea	1.28	-	1.37	-	1.39
ali-46	Bis-(2,2,2-trinitroethyl)-succinate	1.48	-	1.54	-	1.61
ali-47	Bis-(2,2-dinitropropyl)-oxalate	2.36	-	2.33	-	1.96
ali-48	N,N'-Bis-(3,3,3-trinitro-propyl)-oxamide	1.65	-	1.59	-	1.58
ali-49	2,2-Dinitrobutyl-4,4,4-trinitrobutyramide	2.00	-	2.15	-	1.95
ali-50	Bis-(2,2,2-trinitroethyl)-4,4-dinitroethyl-4,4-dinitroheptanedioate	1.83	-	1.66	-	1.58
aro-35	1,2,4,5-Tetranitrobenzene	1.43	-	-	1.47	1.43
aro-36	2,3,4,5,6-Pentanitroaniline	1.18	-	-	1.32	1.31
aro-37	2,4,6-Trinitroaniline	2.25	-	-	2.14	1.86
aro-38	1,3-Diamino-2,4,6-trinitrobenzene	2.51	-	-	2.55	1.99
aro-39	Ammonium picrate	2.13	-	-	2.39	1.90
aro-40	2,4,6-Trinitrobenzonitrile	2.15	-	-	-9.67	1.93
aro-41	1,4,5,8-Tetranitronaphthalene	2.00	-	-	0.69	1.92
aro-42	2',2'-Dinitropropyl-2,4,6-trinitrobenzoate	2.33	-	-	1.57	1.61
aro-43	2,2',4,4',6,6'-Hexanitrodiphenylamine	1.68	-	-	1.66	1.60
aro-44	2,2'',4,4',4'',6,6',6''-Octanitro-m-terphenyl	1.80	-	-	1.68	1.99
aro-45	2,2',4,4',6,64-Hexanitrobibenzyl	2.06	-	-	1.75	1.55
aro-46	2,2',2'',4,4',4'',6,6',6''-Nonanitro-m-terphenyl	1.59	-	-	1.70	1.92
aro-47	2,2',2'',4,4'',5,6,6''-Octanitro-p-terphenyl	1.60	-	-	1.71	1.71
aro-48	2,4,6-Trinitrobenzaldoxime	1.62	-	-	-0.59	1.99
aro-49	1-(3,3,3-Trinitropropyl)-2,4-dinitrobenzene	1.49	-	-	1.74	1.46

| aro-50 | 3-Methyl-2,2',4,4',6,6'-hexanitrobiphenyl | 1.72 | - | - | 1.81 | 1.81 |
| aro-51 | 2,2',4,4',6,6'-Hexanitrobiphenyl | 1.93 | - | - | 1.73 | 1.88 |

Table 2 – Parameters used to evaluate the performances of QSPR models

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$	• y_i : observed property • $\hat{y}_i$: calculated property • $\bar{y}$: mean value of property
$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n - p - 1}}$$	• n : number of compounds • p : number of descriptors
$$t = \frac{b_i}{s_{b_i}}$$	• b_i : regression constant of the descriptor • $s_{b_i} = \dfrac{s}{\sqrt{\sum_i (x_i - \bar{x})^2}}$: standard deviation of b_i • $\bar{x}$: mean value of descriptor
$$Q^2 = 1 - \frac{\sum_i (y_i - \hat{y}_{i/i})^2}{\sum_i (y_i - \bar{y})^2}$$	• $\hat{y}_{i/i}$: predicted property of the ith compound from the model derived without the ith compound in the cross-validation procedure
$$Q^2_{EXT} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_{TR})^2}$$	• $\bar{y}_{TR}$: mean value of property in the training set

Table 3 – Performances of QSPR models

	Nitramines Eq. 2	Nitroaliphatics Eq. 3	Nitroaromatics Eq. 4	Global Eq. 5
n_{TR}	40	34	34	108
n_{EXT}	20	16	17	53
R^2	0.919	0.929	0.812	0.816
$RMSE_{TR}$	0.11	0.13	0.16	0.19
Q^2_{LOO}	0.888	0.901	0.751	0.793
Q^2_{10CV}	0.878	0.893	0.743	0.786
Q^2_{5CV}	0.867	0.902	0.751	0.789
R^2_{YS}	0.103	0.123	0.127	0.049
SD_{YS}	0.071	0.076	0.082	0.028
R^2_{EXT}	0.878	0.876	0.013	0.759
Q^2_{EXT}	0.853	0.863	-71.09[a]	0.749
$RMSE_{EXT}$	0.16	0.19	3.50	0.23
R^2_{IN}	0.845	0.877	0.541	0.750
Q^2_{IN}	0.817	0.871	0.508	0.740
$RMSE_{IN}$	0.16	0.19	0.32	0.22

[a] value out of the classical range due to non physical prediction values obtained for compounds that were finally out of the applicability domain of the model

Table 4 – Comparison of the performances of previous and new MLR-based QSPR models

References	Target compounds	n_{train}	n_{valid}	R^2	Q^2_{LOO}	R^2_{EXT}	Q^2_{EXT}	$RMSE_{EXT}$
Wang et al.[b]	Nitramines	45	10	0.852	0.695	0.856	0.670	0.23
[38]	Nitroaliphatics	43	9	0.801	0.512	0.932	0.811	0.19
	Nitroaromatics	39	10	0.840	0.529	0.457	0.522	0.27
	Nitro compounds	127	29	0.771	0.593	0.715	0.716	0.25
Present work[c]	Nitramines	40	20	0.919	0.888	0.878	0.853	0.16
	Nitroaliphatics	34	16	0.929	0.901	0.876	0.863	0.19
	Nitroaromatics	34	17	0.812	0.751	0.013	-71.09[a]	3.50
	Nitro compounds	108	53	0.816	0.793	0.759	0.749	0.23

[a] Value out of the classical range due to non physical prediction values obtained for compounds that were finally out of the applicability domain of the model

[b] Experimental values collected from various literature references, including ref [39].

[c] All experimental values collected from ref [39].

Table 5 – Predictive powers in the applicability domain (R^2_{in}) of the new QSPR models by family of nitro compounds

Models		Nitramines	Nitroaliphatics	Nitroaromatics	All
Local	Eq. 2	0.845	-	-	0.829
	Eq. 3	-	0.877	-	
	Eq. 4	-	-	0.541	
Global	Eq. 5	0.835	0.816	0.227	0.750

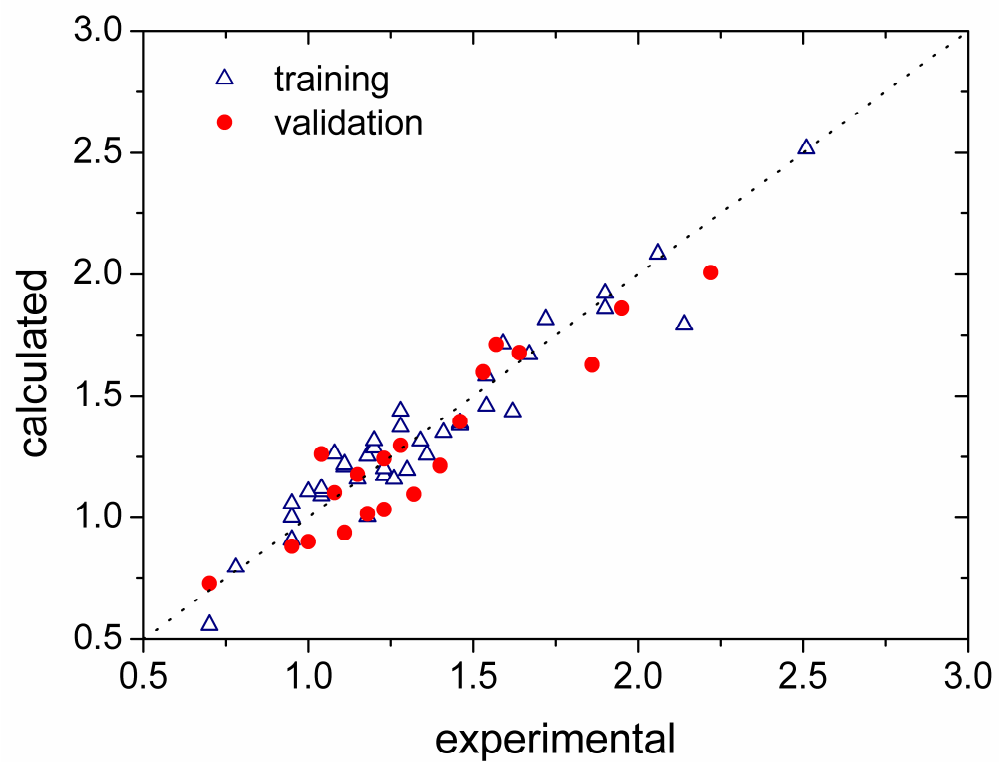**Figure 1** – Experimental vs. predicted $logh_{50\%}$ of nitramine compounds based on Eq. 2

Figure 2 – Experimental vs. predicted $\log h_{50\%}$ of nitroaliphatic compounds based on Eq. 3
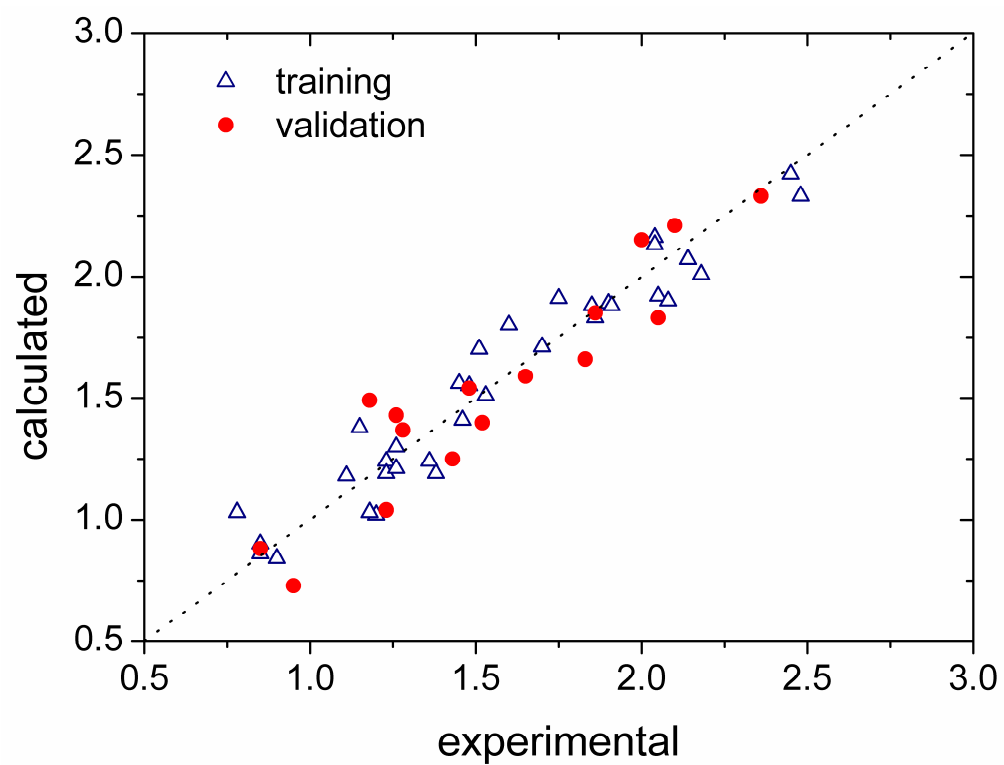
Figure 3 – Experimental vs. predicted $logh_{50\%}$ of nitroaromatic compounds based on Eq. 4
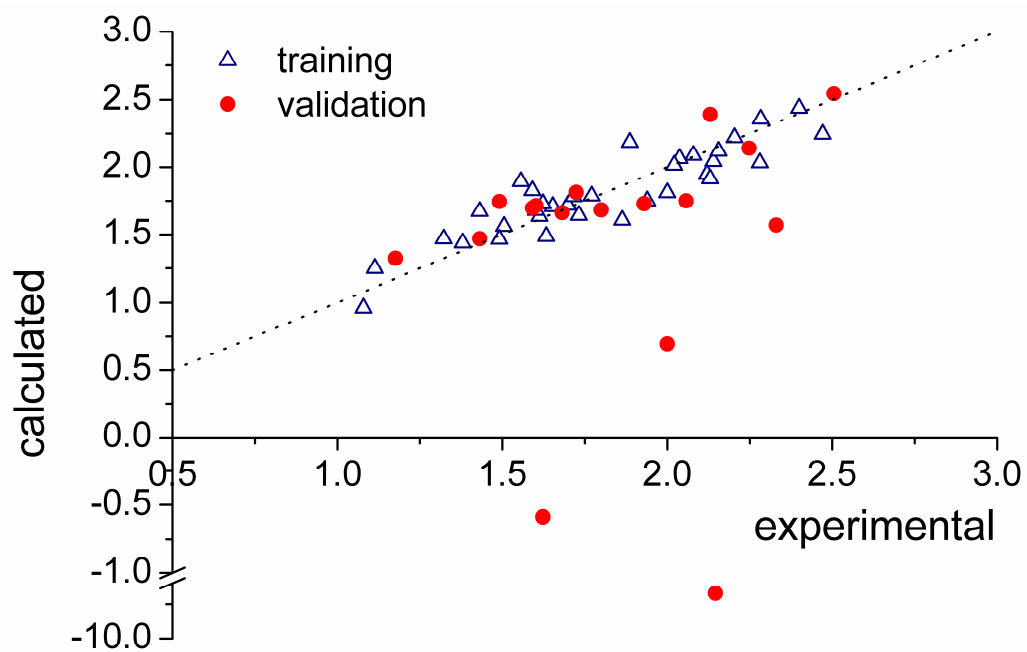
Figure 4 – Identification of clusters of compounds from the score plot of PLS analysis on the 51 nitroaromatic compounds of the data set based on 313 molecular descriptors. The intercalation fragments "A" identified in the data set are: -NH-, -CH$_2$-CH$_2$-, -CH=CH-, -(C=O)- or absent (no intercalation fragment).
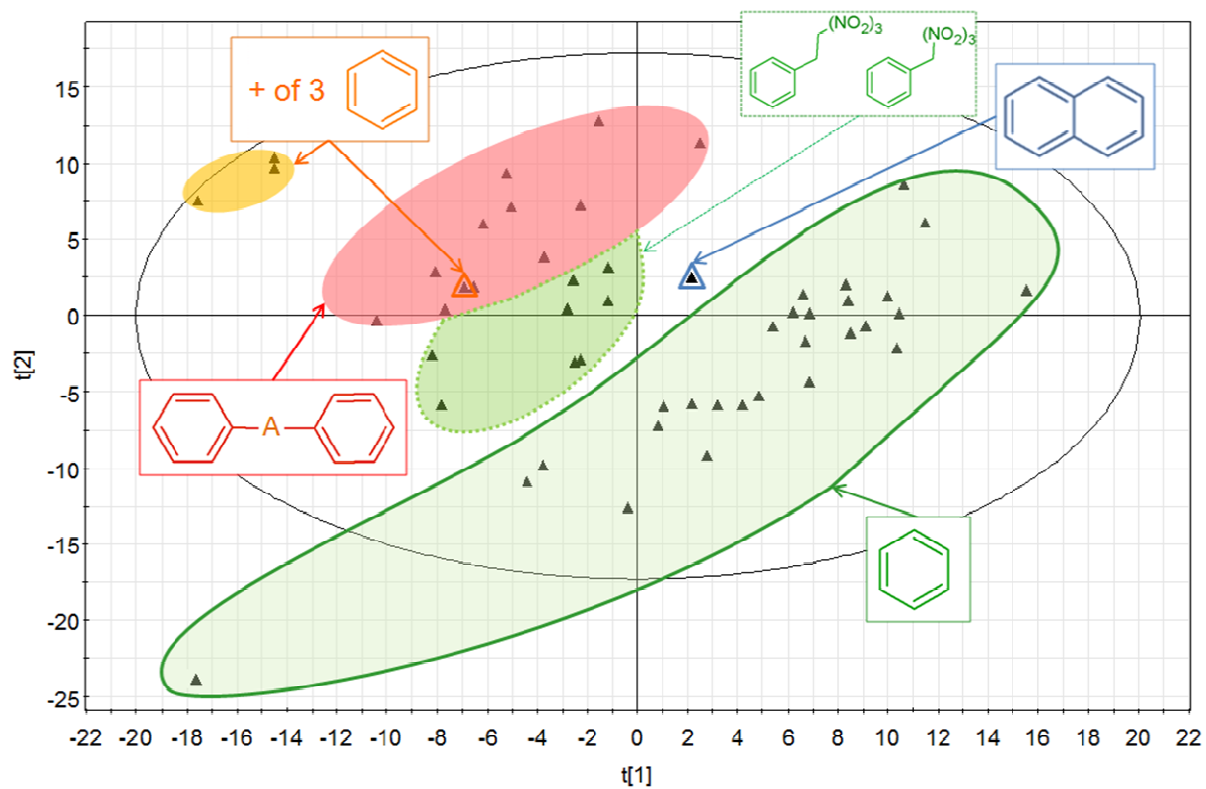
Figure 5 – Experimental vs. predicted $logh_{50\%}$ of nitro compounds based on Eq. 5