J Mol Model (2011) 17:2443–2453
DOI 10.1007/s00894-010-0908-0

ORIGINAL PAPER

Development of a QSPR model for predicting thermal stabilities of nitroaromatic compounds taking into account their decomposition mechanisms

Guillaume Fayet · Patricia Rotureau · Laurent Joubert · Carlo Adamo

Received: 14 September 2010 / Accepted: 16 November 2010 / Published online: 21 December 2010

Abstract The molecular structures of 77 nitroaromatic compounds have been correlated to their thermal stabilities by combining the quantitative structure–property relationship (QSPR) method with density functional theory (DFT). More than 300 descriptors (constitutional, topological, geometrical and quantum chemical) have been calculated, and multilinear regressions have been performed to find accurate quantitative relationships with experimental heats of decomposition ($-\Delta H$). In particular, this work demonstrates the importance of accounting for chemical mechanisms during the selection of an adequate experimental data set. A reliable QSPR model that presents a strong correlation with experimental data for both the training and the validation molecular sets (R^2=0.90 and 0.84, respectively) was developed for non-ortho-substituted nitroaromatic compounds. Moreover, its applicability domain was determined, and the model's predictivity reached 0.86 within this applicability domain. To our knowledge, this work has produced the first QSPR model, developed according to the OECD principles of regulatory acceptability, for predicting the thermal stabilities of energetic compounds.

Keywords Quantitative structure–property relationship (QSPR) · Nitroaromatic compounds · Heat of decomposition · REACH regulation · Density functional theory (DFT)

G. Fayet (✉) · P. Rotureau
Institut National de l'Environnement Industriel et des Risques (INERIS), Parc Technologique Alata, BP2,
60550, Verneuil-en-Halatte, France
e-mail: guillaume.fayet@ineris.fr

G. Fayet · C. Adamo
Laboratoire d'Electrochimie,
Chimie des Interfaces et Modélisation pour l'Energie,
CNRS UMR-7575,
Ecole nationale Supérieure de Chimie de Paris - Chimie ParisTech,
11 rue P. et M. Curie,
75231, Paris Cedex 05, France

L. Joubert
Institut de Recherche en Chimie Organique Fine (IRCOF),
Université de Rouen,
76821, Mont Saint-Aignan, France

Introduction

After initiation by various external stimuli (impact, electric discharge, heat), energetic materials can undergo decomposition reactions, leading to the release of large amount of energy. Therefore, it is very important to understand the thermal stability properties of materials, as we can then evaluate not only their explosive power, but also any hazards associated with their storage, transportation and handling [1]. Indeed, many reported (industrial) accidents have been caused by a lack of knowledge of decomposition processes [2], particularly those concerning nitro compounds [3].

In order to evaluate the amount of energy released during their decomposition, calorimetric analyses are performed on chemicals. Differential scanning calorimetry (DSC) [1, 4–6], which provides heats of decomposition with errors of about 5–10% [7], is a typical example of an experimental screening test.

Such measurements represent a way to preselect substances that may have explosive properties according to the *Recommendations on the Transport of Dangerous Goods* [8], which are used in REACH (registration, evaluation, authorisation and restriction of chemicals) [9] and CLP

(classification, labeling and packaging of substances and mixtures) [10] regulations for classification purposes. Within this new European regulatory framework dedicated to chemicals, the evaluation of a tremendous number of substances may be required in a restricted timeframe. The complete experimental characterization of chemicals causes not only time, cost and ethical (e.g., animal testing in toxicology and ecotoxicology) problems, but can also be hazardous in the case of potentially explosive compounds, (e.g., nitroaromatic compounds) [11]. In this context, the development of alternative tools, notably those deriving from computational chemistry [12], may be useful, and is even recommended in these regulations for the screening and prioritization of chemicals for experiments.

Among the methods available, quantitative structure–property relationships (QSPRs) represent powerful tools for predicting different properties of chemicals. Widely used in biology [13, 14], toxicology [15, 16] and drug design [17, 18], they have been increasingly applied to physicochemical properties for many years [19, 20], particularly the properties of energetic materials [21–32]. The principle of QSPRs is to develop a mathematical relationship connecting a macroscopic property of a series of compounds to microscopic descriptors derived from their molecular structures, using a reliable experimental data set. Once a model has been developed and validated, it can be used to predict the values of that macroscopic property for other molecules with similar structures that have not yet been characterized and may not even have been synthesized. Moreover, it may help us to understand the investigated phenomena at the molecular scale. To encourage the development and use of QSAR/QSPR models, the Organization for Economic Co-operation and Development (OECD) recently introduced five principles for the validation of these models for regulatory purposes [33]. They require a defined endpoint, an unambiguous algorithm, a defined domain of applicability, appropriate measures of goodness-of-fit, robustness and predictivity, and a mechanistic interpretation (if possible). Until now, only a few QSPR models have been investigated based on these principles [34].

Some works have been devoted to predicting the thermal stabilities of chromophores [35], polymers [36], and ionic liquids [37]. Within the framework of energetic materials, some links between the presence of particular functional groups and the decomposition temperatures (T_{onset}) of potentially explosive compounds were highlighted by Grewer in the early 1990s [38]. The popular CHETAH software, used for the prediction of reactivity hazards, estimates the maximum heats of decomposition of chemicals using Benson's group contribution method [39]. Nevertheless, this quantity cannot be directly compared to experimental heats of decomposition since it considers the maximal decomposition of the molecule. More recently, T_{onset} was correlated to the dissociation energies of the weak bonds in nitro molecules [27, 40].

To the best of our knowledge, the first QSPR-type analysis related to the thermal stabilities of nitroaromatic compounds was realized by Saraf [26], based on 19 differential scanning calorimetric data. This study proposed that the number of nitro groups (n_{NO2}) alone could be used to estimate the decomposition enthalpy ($-\Delta H$) with an average error about 8%, a level of uncertainty that is close to the experimental error. More recently, Keshavarz proposed two QSPR models based on constitutional descriptors that could be used to predict the activation energy for the thermolysis of nitroaromatics [24] and nitramines [23]. R^2 was 0.87 for the correlation in both models.

In previous works, we proposed preliminary QSPR models for predicting the heats of decomposition of 22 nitroaromatic compounds [41–43], with the most robust model presenting a good correlation with experimental data (R^2=0.98). Nevertheless, a lack of experimental data (only 22 molecules) meant that it was not possible to estimate the predictive power and applicability domains of the models.

This paper deals with the application of an original approach combining the QSPR methodology with density functional theory (DFT) in order to predict the heats of decomposition of a large series of nitroaromatic compounds (77 molecules), according to the OECD recommendations for the validation of such predictive models. The chosen quantum chemical level of theory (DFT) allows for the optimization of molecular structures and the accurate calculation of a series of chemically comprehensive descriptors of molecular reactivity (e.g., conceptual DFT descriptors) [41–43]. Furthermore, the importance of training set selection is demonstrated, and it will be defined so as to account for subjacent microscopic decomposition mechanisms.

Materials and methods

Experimental data set

The choice of the experimental data set is, of course, a critical point of any QSPR analysis. Since experimental conditions can exert a strong influence, all experimental values used in the fitting procedure should be obtained under the same conditions to ensure their reliability and compatibility.

In this study, a data set of 77 nitroaromatic compounds was considered. It consists of nitrobenzene derivatives, including mono-, di- and trinitrobenzenes and a large variety of substituents (e.g., nitro groups, carboxylic acids or halogens). All of the heats of decomposition (shown in

Tables 1 and 2) were extracted from a single reference [7] to ensure that they were obtained using a single protocol. In the present case, a pressure DSC apparatus was used at a heating rate of 10 $K\,min^{-1}$ for 1–2 mg samples in aluminum cells with a pinhole (i.e., open sample cells).

Molecular structures

All molecular structures were calculated using density functional theory (DFT) with the Gaussian 03 package [44]. Geometry optimizations were performed using the parameter-free PBE0 hybrid functional [45] and the 6-31+G (d,p) basis set. Vibrational frequencies were computed at the same theoretical level to ensure that all stable species corresponded to energy minima. These molecular structures were then loaded into CodessaPro software [46], including information about geometry, atomic charges, molecular orbital energies and vibrational frequencies. More than 300 descriptors were calculated, and some particular external descriptors were also included, as their utility was shown in previous work [41, 42]. This was notably the case for the number of nitro groups and some conceptual DFT descriptors that have already been correlated successfully with thermal stabilities of similar nitroaromatic compounds [41, 42].

The molecular descriptors considered here can be divided into several classes: constitutional, topological, geometric, and quantum chemical. Constitutional descriptors characterize the presence and number of specific atoms, groups or bonds in the molecule (e.g., number of O atoms, single bonds). Topological indices, like the Wiener index, are based on atomic connectivity, and provide information about the size and degree of branching of molecules. Geometric descriptors, contrary to the previous classes calculated from 2D structures, stem from the 3D structure (e.g., molecular volume). Quantum chemical descriptors combine binding, formation, molecular orbital energies, thermodynamic and electronic information, such as the dipole moment or polarizability,with descriptors characterizing the charge distribution within the molecule (e.g., partial charges). Detailed definitions of and information about descriptors can be found in [47].

Statistical analysis

Various data-mining tools can be used to develop QSPR models, such as artificial neural networks [48] and genetic algorithms [49]. In this study, multilinear regressions were computed, and the final model has the following general mathematic form:

$$Y = a_0 + \sum_{i=1}^{n} a_i X_i, \qquad (1)$$

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

Equation 1 was obtained using the "best multilinear regression" (BMLR) technique, described in [47] and implemented with the CodessaPro software.

The first step in the BMLR analysis is to reduce the initial set of descriptors by rejecting all descriptors with insignificant variance, thus ensuring that descriptors are not included by chance in the model when they are not related to the investigated property. At the same time, if two descriptors are highly intercorrelated, only the one that has the strongest correlation with the property is retained. This step not only prevents the introduction of inappropriate descriptors, but it also makes the analysis faster, as fewer variables must be treated.

After that, starting from pairs of orthogonal (i.e., not intercorrelated) descriptors, higher rank models are computed by successively including orthogonal descriptors as soon as an increase in correlation is observed. Then, the BMLR analysis selects the best models at each rank, and the final model must be chosen from among them. This has to be sufficiently correlated and, at the same time, protect against any overparameterization, which would lead to a loss of predictive power for molecules outside the training set.

In this work, the "breaking point" rule was used to manage this problem. This method, which has already been successfully used in previous works [22, 43, 50], consists of analyzing the improvement in the correlation with the number of variables in the model. By plotting the R^2 values as functions of the number of descriptors, asymptotic behavior was observed, and the improvement in the correlation became less significant after a certain rank ($\Delta R^2 < 0.02$–0.03). At this point (the "breaking point"), the model is considered to be optimal, representing the best compromise between correlation and parameterization.

The robustness and the stability of the models were evaluated through the square of the correlation coefficient (R^2) and the mean absolute error (MAE). The choice of the descriptors was confirmed by performing Student's t-test at a confidence level of 95%. Moreover, they were validated internally using the cross-validation technique (R^2_{cv}), and externally, using a validation data set (R^2_{valid}).

Finally, the applicability domain of the models (i.e., the domain in which predictions are reliable) was investigated. In the present paper, the applicability domain estimation was performed based on Euclidean distances in the descriptor space (after pre-processing with a principal component analysis) using Ambit Discovery software [51]. The applicability domain was defined to include 95% of the training set molecules, and all predictions within this domain are expected to be reliable. This last

Table 1 Experimental and calculated heats of decomposition ($-\Delta H$) in kJ mol^{-1} for the molecules of the training set

Molecules	Experimental values [7]	Calculated via Eq. 2	Calculated via Eq. 3	Calculated via Eq. 4
Non-ortho compounds				
2-Amino-4-nitrophenol	130	238	173	-
3-Nitrotoluene	149	238	212	-
2-Amino-5-nitrophenol	153	201	239	-
4-Nitrotoluene	213	253	192	-
4-Nitrophenol	232	183	235	-
3-Nitroanisole	243	223	288	-
3-Nitrobenzoic acid methyl ester	256	334	277	-
2,6-Dichloro-4-nitroaniline	264	394	284	-
4-Nitrophenetole	270	347	249	-
4-Nitrophenylhydrazine	277	389	279	-
3-Nitrophenol	283	165	227	-
3-Nitrobenzoic acid	289	267	372	-
4-Nitroacetophenone	291	348	343	-
4-Nitrobenzyl alcohol	292	223	272	-
4-Nitrobenzoic acid methyl ester	302	329	264	-
4-Nitro-2-toluidine	306	287	315	-
4-Nitrobenzamide	319	275	321	-
4-Nitrobenzyl chloride	337	675	333	-
4-Nitroaniline	347	298	308	-
3-Nitroaniline	350	302	317	-
3-Nitrophenylacetic acid	358	397	347	-
4-Nitrobenzhydrazide	362	415	335	-
3-Nitroacetoanilide	369	289	394	-
2-Amino-4-nitroanisole	375	339	325	-
4-Nitroacetoanilide	387	343	372	-
4-Nitrobenzoyl chloride	408	463	303	-
3-Nitrocinnamic acid	414	314	417	-
4-Nitrobenzaldehyde	421	495	394	-
3,5-Dinitrobenzonitrile	654	699	698	-
3,5-Dinitrobenzoic acid	674	658	679	-
3,5-Dinitrobenzylchloride	711	682	673	-
Ortho compounds				
5-Chloro-2-nitrobenzotrifluoride	40	96	-	7
2-Nitrophenol	123	176	-	172
2-Nitrophenylacetic acid	175	288	-	266
2-Nitrotoluene	182	223	-	211
2-Nitroanisole	230	203	-	276
2-Nitrobenzamide	256	317	-	341
2-Nitrobenzoic acid	271	212	-	247
2-Nitrobenzoic acid methyl ester	274	284	-	285
2-Nitroaniline	280	267	-	284
2-Nitroacetophenone	308	243	-	244
2-Nitrobenzaldehyde	318	196	-	351
2-Nitrobenzyl alcohol	319	350	-	239
1-Chloro-3,4-dinitrobenzene	342	358	-	331
2-Nitrophenylsulfenyl chloride	344	302	-	276
4-Chloro-2-nitroaniline	349	263	-	312

Table 1 (continued)

Molecules	Experimental values [7]	Calculated via Eq.2	Calculated via Eq.3	Calculated via Eq.4
2-Nitrophenylhydrazine	381	448	-	298
1,5-Dinitro-2,4-difluorobenzene	439	396	-	483
5-Nitrovanillin	450	448	-	491
2-Nitro-5-thiocyanotobenzoic acid	492	718	-	559
2,4-Dinitroaniline	597	520	-	694
2-Nitrobenzhydrazide	598	550	-	463
2-Nitrocinnamic acid	600	620	-	606
3,4-Dinitrotoluene	684	645	-	626
3,4-Dinitrobenzoic acid	701	706	-	803
2,6-Dinitroaniline	719	713	-	619
2-Chloro-3,5-dinitrobenzoic acid	1023	766	-	937
2,4,6-Trinitrotoluene	1223	1149	-	1220
	R^2	0.84	0.90	0.94
	MAE (%)	23	12	17
	R^2_{cv}	0.81	0.86	0.91

statement was checked by calculating the correlation coefficients (R^2_{in}) for the molecules of the validation set inside this applicability domain.

Results and discussion

Complete data set

In a first step, the whole data set of 77 molecules was investigated. This data set was sufficiently large that we could divide it into a training and a validation set. In order to retain similar distributions for the two ensembles, the entire set was sorted from the smallest to the highest values of heat of decomposition, and the validation set was generated from the third, seventh, etc. molecules. Thus, the distributions of both the training and the validation sets (58 and 19 molecules, respectively) were quite similar to that of the entire data set, as shown in Fig. 1. This ratio enables both sets to be of a sufficient size to allow the robust development and validation of the model. The BMLR method was then applied to the training set (see Table 1). Models including up to 19 descriptors were built. The improvement in R^2 with the number of descriptors included in the model is shown in Fig. 2. Based on the "breaking point" rule, the following four-parameter model was considered the best compromise between correlation and number of descriptors:

$$-\Delta H = -282.3 + 333.5\, n_{NO2} - 1214.5\, E_{C,avg} + 7.4\, \alpha - 275.6\, {}^0IC_{avg}, \quad (2)$$

where n_{NO2} is the number of nitro groups, $E_{C,avg}$ is the average 1-electron reactivity index for a C atom, α is the mean polarizability, and ${}^0IC_{avg}$ is the average information content (order 0), a topological index. This model has the advantage of including descriptors that are directly related to thermal stability. In fact, the amount of energy released during decomposition is expected to be linked to the loss of nitro groups [52]. It is thus pertinent to find the number of nitro groups in the model. Moreover, $E_{C,avg}$ characterizes the reactivity of carbon atoms in the molecule, and the one connected to the leaving nitro group is expected to be the most reactive in the molecule. Besides, nitro groups also influence the electronic properties of the molecule, notably the polarizability.

This model is significantly correlated with R^2=0.84 and quite stable ($R^2_{cv} = 0.81$). While it is less correlated than our previous work based on 22 molecules (R^2=0.98) [43], the distribution of experimental data in that work was not as homogeneous as that considered in the present work, and the single trinitro molecule of the previous set strongly influenced the regression.

To evaluate the predictive power of this new model, heats of decomposition of the validation set were calculated. As shown in Fig. 3 and Table 2, the model gives only low correlation, with an average deviation of 32% from experimental data ($R^2_{valid} = 0.43$). In particular, the calculated values for 4-fluoro-2-nitrotoluene, 2,4- and 2,6-dinitrobenzoic acids present large errors: 64, 83 and 206%, respectively (see Table 2). Nevertheless, no molecule of this validation set was excluded from the applicability domain of the model (see Table 3). Therefore, the predictive power of the model was low within its own applicability domain ($R^2_{in} = 0.43$).

Table 2 Experimental and calculated heats of decomposition ($-\Delta H$) in kJ mol^{-1} for the molecules of the validation set

molecules	Experimental values [7]	Calculated via Eq. 2	Calculated via Eq. 3	Calculated via Eq. 4	From Saraf [26]	From previous works [43]
Non-ortho compounds						
Nitrobenzene	161	188	202	-	314	319
4-Nitroanisole	248	248	283	-	314	312
4-Nitrophenylacetic acid	265	291	341	-	314	314
3-Nitroacetophenone	276	338	364	-	314	308
4-Nitrobenzoic acid	284	275	332	-	314	267
3-Nitrobenzamide	311	298	334	-	314	39
3-Nitrobenzyl alcohol	325	259	258	-	314	352
3-Nitrobenzaldehyde	373	297	389	-	314	339
3-Nitrobenzhydrazide	430	722	344	-	627	631
4-Nitrocinnamic acid	506	355	414	-	314	413
3,5-Dinitrobenzamide	736	721	687	-	627	560
Ortho compounds						
4-fluoro-2-Nitrotoluene	129	212	-	211	314	354
2,6-Dinitrobenzoic acid	222	681	-	423	627	607
2-Nitroacetoanilide	297	431	-	472	314	80
4-Nitro-3-cresol	345	204	-	203	314	319
4-Chloro-3-nitrobenzoic acid	354	238	-	463	314	325
2,4-Dinitrobenzoic acid	394	306	-	553	314	280
2,4-Dinitrotoluene	632	585	-	607	314	205
3,4-Dinitrobenzylalcohol	685	660	-	683	627	605
	R^2_{valid}	0.43	0.84	0.42	0.53	0.37
	MAE (%)	32	18	44	36	46

At this point, standard experimental uncertainty is not sufficient to explain the poor predictive power of the model. Another factor affecting the accuracy of the model is related to the chemical reactivity. Indeed, the decomposition of nitroaromatic compounds is a complex process. Whereas the direct breaking of the carbon nitro bond was sometimes considered the initiation step in the decomposition of nitro compounds [40], numerous studies have demonstrated that more complex reaction paths can be involved, particularly in nitroaromatic molecules [52] such as ortho-substituted nitrobenzene derivatives (e.g., *o*-nitrotoluenes [53] and their derivatives [54], including 2,4,6-trinitrotoluene [55]).

For this reason, models have been developed for non-ortho and ortho nitroaromatic compounds separately; in

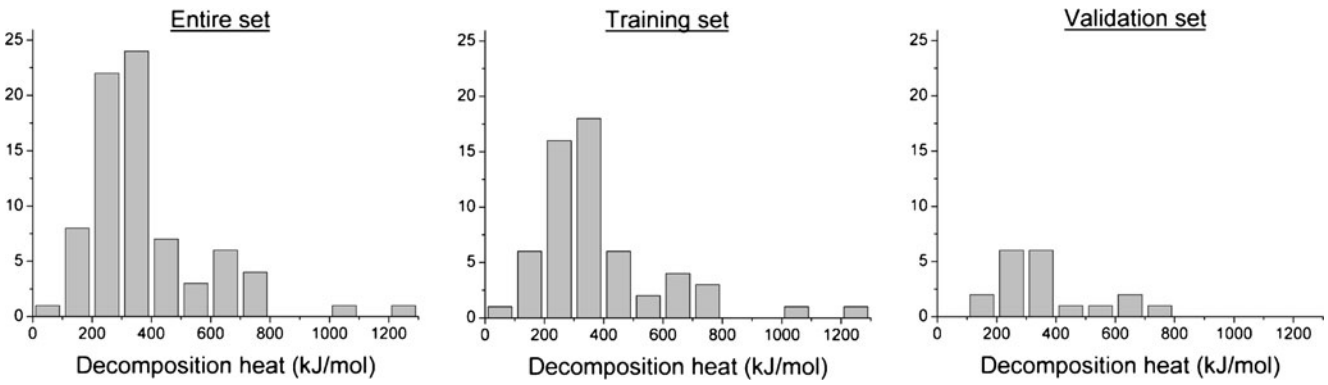

Fig. 1 Distributions of the experimental heat of decomposition in the entire, training and validation sets

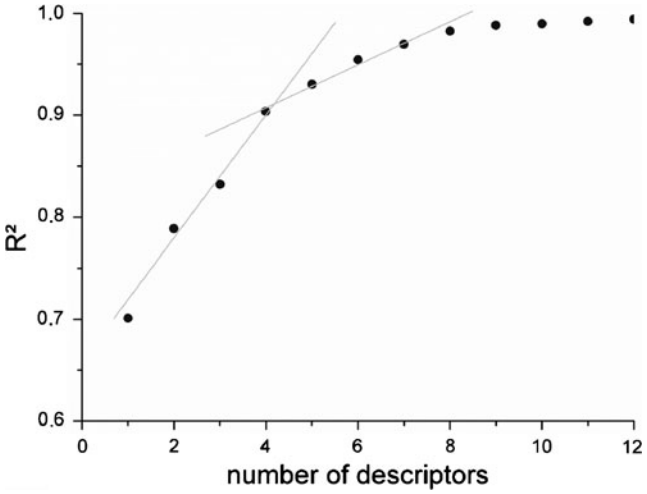

Fig. 2 Number of descriptors versus R^2 of the models from the BMLR analysis of the entire training set

other words, a first model is applicable to compounds that present no substituent at the ortho position to the nitro group, and a second one is applicable to compounds that do present such a substituent. Therefore, two models were developed according to the protocol that was previously used for the entire set, so they were also divided into training and validation sets (in Tables 1 and 2). Storm [56] and Kamlet [57] have already considered the importance of the substituents at this position by pointing out nitroaromatic explosives with an alpha C–H linkage when reporting experimental impact sensitivities and their correlation with the oxygen balance.

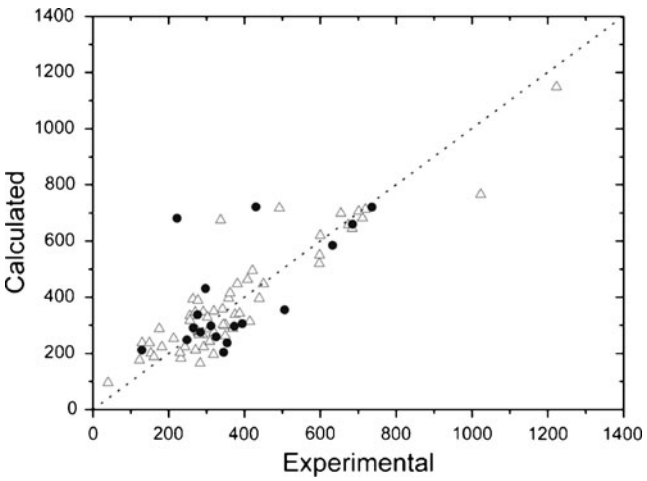

Fig. 3 Experimental versus calculated heats of decomposition (in kJ mol^{-1}) of nitroaromatic compounds according to Eq. 2 (the molecules of the training and validation sets are represented by *wide triangles* and *plain circles*, respectively)

Non-ortho compounds

A four-parameter model was developed from the 31 non-ortho molecules of the training set:

$$-\Delta H = 0.8\,G - 3.8\,\mathrm{WPSA1} - 4255.1\,Q_{\max} + 26.8\,\mathrm{RPCS} - 251.2, \quad (3)$$

where G is the gravitational index, WPSA1 is the weighted positive surface area (obtained via calculated Mulliken charges [58]), $Q_{\max}$ is the maximal partial charge on the molecule (calculated according to Gasteiger's method [59]), and RPCS is the relative positively charged surface area (obtained via calculated Zefirov charges [47]). While $Q_{\max}$ is related to the nitrogen atoms in the nitro groups, the other descriptors are more difficult to link to the decomposition process. G characterizes the molecular shape and the mass distribution of the molecule. WPSA1 and RPCS are charged partial surface area descriptors, developed to encode the features responsible for the polar interactions between molecules. Nevertheless, these last descriptors are related, in a certain way, to the distribution of charge within the molecule, which is very influenced by nitro groups. Therefore, they are indirectly related to the properties of this group, which is central to the decomposition of nitroaromatic compounds.

This model is strongly correlated with experimental data (R^2=0.90), with an average deviation of 12%, close to the experimental uncertainty (see Fig. 4 and Table 1). The internal validation, performed by cross-validation, gives a satisfactory result: $R^2_{cv} = 0.86$. Moreover, the model presents good predictivity, as calculated values are close to experimental ones for the 11 non-ortho molecules of the validation set ($R^2_{valid} = 0.84$, see Table 2). Regarding the experimental uncertainty, the performance of this model is satisfactory, as the predictive power of the model is as high as 0.86 when the molecule that was found to be outside the applicability domain of the model (see Table 3) is excluded.

Ortho compounds

For ortho nitroaromatic compounds, the final model, developed using the data for the 27 molecules of the training set, is another four-parameter equation:

$$-\Delta H = 4.1\,\mathrm{PNSA1} - 3298.8\,\mathrm{RPCG} - 56228\,N_{\mathrm{C,min}} - 1245.5 S_{\mathrm{ZX/ZX}} + 1117.8, \quad (4)$$

where PNSA1 and RPCG are the partial negative charged surface area and the relative positive charge (obtained via calculated Zefirov charges), $N_{\mathrm{C,min}}$ is the minimum nucleophilic reactivity index for a C atom, and $S_{\mathrm{ZX/ZX}}$ (ZX

Table 3 Analysis of the applicability domain of QSPR models

	Training set					Validation set				
	R^2	N_{in}	N_{out}	$\%_{in}$	R^2_{in}	R^2_{valid}	N_{in}	N_{out}	$\%_{in}$	R^2_{in}
Calculated via Eq. 2	0.84	55	3	95	0.82	0.43	19	0	100	0.43
Calculated via Eq. 3	0.90	29	2	94	0.88	0.84	10	1	90	0.86
Calculated via Eq. 4	0.94	26	1	96	0.94	0.42	8	0	100	0.42
From Saraf [26]	0.98	19	0	100	0.98	0.53	17	2	89	0.54
From previous works [43]	0.98	21	1	95	0.99	0.37	15	4	79	0.24

N_{in} is the number of molecules in the applicability domain

N_{out} is the number of molecules outside the applicability domain

$\%_{in}$ is the percentage of molecules in the applicability domain

R^2_{in} is the predictive power in the applicability domain

shadow / ZX rectangle) characterizes the molecular shape within the *ZX* plane. The charge-related descriptors are the main descriptors in this equation. The only descriptor that is directly related to the thermal stability is $N_{C,min}$, which is influenced by the electronic properties of nitro groups, and thus to their reactivity with the aromatic ring.

Concerning the performance of the model, the correlation with the experimental data is high (R^2=0.94), with an average deviation of 17%, and the cross-validation procedure exhibits good robustness ($R^2_{cv} = 0.91$), but its predictivity is low as calculated values for the eight ortho molecules of the validation set deviate by about 44% from the experimental data ($R^2_{valid} = 0.42$).

In fact, there was a stronger expectation of a reliable model for the non-ortho than for the ortho molecules, as there are many different reaction paths depending upon the nature of the substituent at the position ortho to the nitro group (nitro, alcohol, amino…) [52], whereas all non-ortho nitroaromatic compounds follow the same decomposition process (C–NO_2 homolysis, as shown in previous theoretical studies [60, 61]) as their main reaction path (without any influence from the nature of the substituents in the positions meta and para to the nitro group). Besides, applicability domain analysis confirms this observation, as the predictive powers (R^2_{in}) of models 2 (entire set) and 4 (ortho molecules) are low in their applicability domains compared to that of model 3 (non-ortho molecules).

To compare the models developed in this study to previous ones, the heats of decomposition of the validation molecules were calculated using Saraf's model [26] (based on the number of nitro groups of 19 nitrobenzene derivatives) and our previous model [43] (based on 22 molecules), as shown in Table 2. No satisfactory correlation was seen, as the average deviations were 36% and 46% (R^2 =0.53 and 0.37) for Saraf's model and our previous work, respectively. This may be due to the lack of homogeneity in the data sets, which also consider no or only one trinitro compound. Besides, these models did not distinguish ortho and non-ortho compounds. Finally, the predictive powers of these models are very low, even in their respective applicability domains (as shown in Table 3). These previous models present the same limit as Eqs. 2 and 4 by considering different decomposition mechanisms in the same data set. Thus, a knowledge of the molecular decomposition path is crucial to the accurate prediction of nitroaromatic thermal stabilities.

Accordance with OECD principles

Our best QSPR model, developed for non-ortho compounds (Eq. 3), follows the five OECD principles for validating QSAR/QSPR models for regulatory use [33]:

Principle 1 The endpoint is well defined as the heat of decomposition, as measured using a pressure

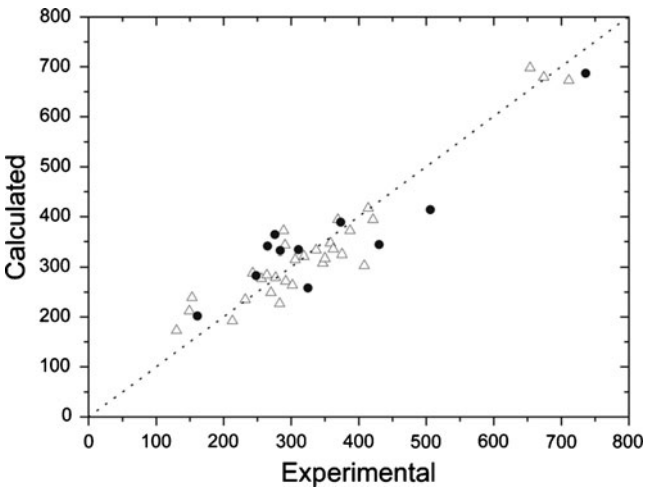

Fig. 4 Experimental versus calculated heats of decomposition (in kJ mol^{-1}) of non-ortho nitroaromatic compounds according to Eq. 3 (the molecules of the training and validation sets are represented by *wide triangles* and *plain circles*, respectively)

DSC apparatus with a heating rate of 10 K min^{-1} on 1–2 mg samples in aluminum cells with a pinhole [7].

Principle 2 The model is transparent, as it consists of a simple multilinear equation, including four parameters, calculated from well-defined DFT-optimized structures at the PBE0/6-31+G(d,p) level.

Principle 3 The model is applicable to all nitrobenzene derivatives without a substituent at the position ortho to the nitro group that are included in an applicability domain defined by interpolating the training set.

Principle 4 The performance of the model is estimated via its correlation with the training set (R^2=0.90), its robustness (as checked by cross-validation: $R^2_{cv} = 0.86$), and its predictive power in its domain of applicability on an external validation set of molecules ($R^2_{in} = 0.86$).

Principle 5 While no mechanistic interpretation can be easily provided from descriptors used in the model, the molecular mechanisms were considered during its development and form part of the defined applicability domain, since the model is not applicable to ortho-substituted compounds (see Principle 3). Indeed, the molecular mechanism involved in the decomposition of the target non-ortho substituted nitroaromatic compounds has been characterized from density functional theory calculations in previous work [57].

As this model satisfies all of the requirements of OECD principles, it could be used as an efficient alternative to experimental characterization as a first screening test to evaluate whether a target nitroaromatic compound may have explosive properties.

Conclusions

A set of 77 nitroaromatic compounds was utilized to develop a reliable QSPR model for predicting their heats of decomposition. Molecular structures calculated at the DFT level of theory were described using more than 300 descriptors.

Considering the entire set of data led to significant correlation in the training step (R^2=0.84), but it failed to predict the heats of decomposition of external molecules (about 32% in deviation). While experimental uncertainty (e.g., the use of open sample cells) contributed to the lack of predictive power, this paper demonstrated that data set selection must also take into account chemical reactivity during the decomposition process.

Indeed, considering a set of nitro compounds without any substituent ortho to the nitro group led to a performant model with significant correlation between calculated and experimental heats of decomposition of molecules from the validation set ($R^2_{valid} = 0.84$). Besides, within its own domain of applicability, the predictivity of this model is very high ($R^2_{in} = 0.86$). The improvement in accuracy compared with the model for the entire set is related to the fact that various specific decomposition mechanisms are involved when substituents are in the position ortho to the nitro groups, whereas all non-ortho compounds decompose along the same reaction path (homolysis of the C–NO_2 bond).

Finally, this model is, to our knowledge, the most reliable QSPR model for predicting the heat of decompositions of nitroaromatic compounds, and the first to be dedicated to energetic compounds while following all OECD requirements for regulatory use.

References

1. Grewer T (1994) Thermal hazards of chemical reactions. Elsevier, Amsterdam
2. Etchells JC (1997) Why reactions run away. Org Process Res Dev 1:435–437
3. Gustin JL (1998) Runaway reaction hazards in processing organic nitro compounds. Org Process Res Dev 2:27–33
4. Jones DEG, Augsten RA (1996) Evaluation of systems for use in DSC measurements on energetic materials. Thermochim Acta 286:355–373
5. Chervin S, Bodman GI (2003) Method for estimating decomposition characteristics of energetic chemicals. Process Saf Prog 22:241–243
6. Yoshida T (1987) Safety of reactive chemicals, industrial safety series, vol 1. Elsevier, Amsterdam
7. Ando T, Fujimoto Y, Morisaki S (1991) Analysis of differential scanning calorimetric data for reactive chemicals. J Hazard Mater 28:251–280
8. United Nations (2009) Recommendations on the transport of dangerous goods: manual of tests and criteria (5th revised edn, ST/SG/AC.10/11/Rev.5). UN, New York
9. EC (2006) Regulation 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). EC, Brussels
10. EC (2008) Regulation N°1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) N° 1907/2006. EC, Brussels
11. Medard L (1987) Les explosifs occasionnels. Tech & Doc, Cachan
12. Lewis A, Kazantzis N, Fishtik I, Wilcox J (2007) Integrating process safety with molecular modeling-based risk assessment of chemicals within the REACH regulatory framework: Benefits and future challenges. J Hazard Mater 142:592–602
13. Gao H, Katzenellenbogen JA, Garg R, Hansch C (1999) Comparative QSAR analysis of estrogen receptor ligands. Chem Rev 99:723–744

14. Winkler DA (2002) The role of quantitative structure–activity relationships (QSAR) in biomolecular discovery. Brief Bioinform 3:73–86
15. Bradbury SP (1995) Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. Toxicol Lett 79:229–237
16. Selassie CD, Garg R, Kapur S, Kurup A, Verma RP, Mekapati SB, Hansch C (2002) Comparative QSAR and the radical toxicity of various functional groups. Chem Rev 102:2585–2606
17. Grover M, Singh B, Bakshi M, Singh S (2000) Quantitative structure–property relationships in pharmaceutical research: Part 2. Pharm Sci Tech Today 3:50–57
18. Grover M, Singh B, Bakshi M, Singh S (2000) Quantitative structure–property relationships in pharmaceutical research: Part 1. Pharm Sci Technol Today 3:28–35
19. Cocchi M, De Benedetti PG, Seeber R, Tassi L, Ulrici A (1999) Development of quantitative structure–property relationships using calculated descriptors for the prediction of the physico-chemical properties (n_D, ρ, bp, ε, η) of a series of organic solvents. J Chem Inf Comput Sci 39:1190–1203
20. Katritzky AR, Lobanov VS, Karelson M (1995) QSPR: the correlation and quantitative prediction of chemical and physical-properties from structure. Chem Soc Rev 24:279–287
21. Katritzky AR, Petrukhin R, Jain R, Karelson M (2001) QSPR analysis of flash points. J Chem Inf Comput Sci 41:1521–1530
22. Katritzky AR, Stoyanova-Slavova IB, Dobchev DA, Karelson M (2007) QSPR modeling of flash points: an update. J Mol Graph Model 26:529–536
23. Keshavarz MH (2009) Simple method for prediction of activation energies of the thermal decomposition of nitramines. J Hazard Mater 162:1557–1562
24. Keshavarz MH, Pouretedal HR, Shokrolahi A, Zali A, Semnani A (2008) Predicting activation energy of thermolysis of polynitro arenes through molecular structure. J Hazard Mater 160:142–147
25. Rice BM, Hare JJ (2002) A quantum mechanical investigation of the relation between impact sensitivity and the charge distribution in energetic molecules. J Phys Chem A 106:1770–1783
26. Saraf SR, Rogers WJ, Mannan MS (2003) Prediction of reactive hazards based on molecular structure. J Hazard Mater 98:15–29
27. Theerlynck E, Mathieu D, Simonetti P (2005) Towards improved models to rationalize and estimate the decomposition temperatures of nitroalkanes, nitramines and nitric esters. Thermochim Acta 426:123–129
28. Afanas'ev GT, Pivina TS, Sukhachev DV (1993) Comparative characteristics of some experimental and computational methods for estimating impact sensitivity parameters of explosives. Propel Explos Pyrotech 18:309–316
29. Toghiani RK, Toghiani H, Maloney SW, Boddu VM (2008) Prediction of physicochemical properties of energetic materials. Fluid Phase Eq 264:86–92
30. Politzer P, Ma Y, Lane P, Concha MC (2005) Computational prediction of standard gas, liquid, and solid-phase heats of formation and heats of vaporization and sublimation. Int J Quantum Chem 106:341–347
31. Politzer P, Martinez J, Murray JS, Concha MC, Toro-Labbe A (2009) An electrostatic interaction correction for improved crystal density prediction. Mol Phys 107:2095–2101
32. Pospisil M, Vavra P, Concha MC, Murray JS, Politzer P (2010) A possible crystal volume factor in the impact sensitivities of some energetic compounds. J Mol Model 16:895–901
33. OECD (2009) Principles for the validation, for regulatory purposes, of (quantitative) structure–activity relationship models. Organisation for Economic Co-operation and Development (OECD), Paris
34. Papa E, Kovarich S, Gramatica P (2009) Development, validation and inspection of the applicability domain of QSPR models for physicochemical properties of polybrominated diphenyl ethers. QSAR Comb Sci 28:790–796
35. Figueiredo LJO, Garrido FMS (2001) Chemometric analysis of nonlinear optical chromophores structure and thermal stability. J Mol Struct THEOCHEM 539:75–81
36. Yu X, Xie Z, Yi B, Wang X, Liu F (2007) Prediction of the thermal decomposition property of polymers using quantum chemical descriptors. Eur Polym J 43:818–823
37. Kroon MC, Buijs W, Peters CJ, Witkamp G-J (2007) Quantum chemical aided prediction of the thermal decomposition mechanisms and temperatures of ionic liquids. Thermochim Acta 465:40–47
38. Grewer T (1991) The influence of chemical structure on exothermic decomposition. Thermochim Acta 187:133–149
39. Benson SW (1976) Thermochemical kinetics, 2nd edn. Wiley, New York
40. Saraf SR, Rogers WJ, Mannan MS (2003) Application of transition state theory for thermal stability prediction. Ind Eng Chem Res 42:1341–1346
41. Fayet G, Joubert L, Rotureau P, Adamo C (2009) On the use of descriptors arising from the conceptual density functional theory for the prediction of chemicals explosibility. Chem Phys Lett 467:407–411
42. Fayet G, Rotureau P, Joubert L, Adamo C (2009) On the prediction of thermal stability of nitroaromatic compounds using quantum chemical calculations. J Hazard Mater 171:845–850
43. Fayet G, Rotureau P, Joubert L, Adamo C (2010) QSPR modeling of thermal stability of nitroaromatic compounds: DFT vs. AM1 calculated descriptors. J Mol Model 16:805–812
44. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03. Gaussian Inc., Wallington
45. Adamo C, Barone V (1999) Toward reliable density functional methods without adjustable parameters: the PBE0 model. J Chem Phys 110:6158–6170
46. University of Florida (2002) CodessaPro. University of Florida, Gainesville (http://www.codessa-pro.com)
47. Karelson M (2000) Molecular descriptors in QSAR/QSPR. Wiley, New York
48. Gasteiger J, Zupan J (1993) Neural networks in chemistry. Angew Chem Int Ed Engl 32:503–527
49. Leardi R (2001) Genetic algorithms in chemometrics and chemistry: a review. J Chemometr 15:559–569
50. Katritzky AR, Dobchev DA, Tulp I, Karelson M, Carlson DA (2006) QSAR study of mosquito repellents using Codessa Pro. Bioorg Med Chem Lett 16:2306–2311
51. Jaworska J, Jeliazkova N (2007) Ambit Discovery 1.20
52. Brill TB, James KJ (1993) Kinetics and mechanisms of thermal decomposition of nitroaromatic explosives. Chem Rev 93:2667–2692
53. Chen SC, Xu SC, Diau E, Lin MC (2006) A computational study on the kinetics and mechanism for the unimolecular decomposition of o-nitrotoluene. J Phys Chem A 110:10130–10134

54. Fayet G, Joubert L, Rotureau P, Adamo C (2009) A theoretical study of the decomposition mechanisms on substituted ortho-nitrotoluenes. J Phys Chem A 113:13621–13627
55. Cohen R, Zeiri Y, Wurzberg E, Kosloff R (2007) Mechanism of thermal unimolecular decomposition of TNT (2, 4, 6-Trinitrotoluene): a DFT study. J Phys Chem A 111:11074–11083
56. Storm CP, Stine JR (1990) Sensitivity relationships in energetic materials. In: Bulusu SN (ed) Chemistry and physics of energetic materials. Kluwer Academic Publishers, Dordrecht, pp 605–630
57. Kamlet MJ, Adolph HG (1979) The relationship of impact sensitivity to structure of organic high explosives. II. Polynitroaromatic explosives. Propel Explos 4:30–34
58. Mulliken RS (1955) Electronic population analysis on LCAO-MO molecular wave functions, I. J Chem Phys 23:1833–1840
59. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36:3219–3228
60. Brill TB, James KJ, Chawla R, Nicol G, Shukla A, Futrell JH (1999) Influence of the substituent on the major decomposition channels of the NO2 group in para-substituted nitrobenzenes: a tandem mass spectrometric study. J Phys Org Chem 12:819–826
61. Fayet G, Joubert L, Rotureau P, Adamo C (2008) Theoretical study of the decomposition reactions in substituted nitrobenzenes. J Phys Chem A 112:4054–4059