HAL
open science

Chemoinformatics for the safety of energetic and reactive materials at Ineris

Guillaume Fayet, Patricia Rotureau

To cite this version:

Guillaume Fayet, Patricia Rotureau. Chemoinformatics for the safety of energetic and reactive materials at Ineris. Molecular Informatics, In press, 10.1002/minf.202000190 . ineris-03318337

HAL Id: ineris-03318337

https://ineris.hal.science/ineris-03318337

Submitted on 9 Aug 2021

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

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

Chemoinformatics for the safety of energetic and reactive materials at Ineris

Guillaume Fayet*, Patricia Rotureau

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

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

Abstract

The characterization of physical hazards of substances is a key information to manage the risks associated to their use, storage and transport. With decades of work in this area, Ineris develops and implements cutting-edge experimental facilities allowing such characterizations at different scales and under various conditions to study all of the dreaded accident scenarios. This review presents the efforts engaged by Ineris more recently in the field of chemoinformatics to develop and use new predictive methods for the anticipation and management of industrials risks associated to energetic and reactive materials as a complement to experiments.

An overview of the methods used for the development of Quantitative Structure-Property Relationships for physical hazards are presented and discussed regarding the specificities associated to this class of properties. A review of models developed at Ineris is also provided from the first tentative models on the explosivity of nitro compounds to the successful application to the flammability of organic mixtures. Then, a discussion is proposed on the use of QSPR models. Good practices for robust use for QSPR models are recalled with specific comments related to physical hazards, notably for regulatory purpose. Dissemination and training efforts engaged by Ineris are also presented. The potential offered by these predictive methods in terms of in *silico design* and for the development of new intrinsically safer technologies in safety-by-design strategies is finally discussed. At last, challenges and perspectives to extend the application of chemoinformatics in the field of safety and in particular for the physical hazards of energetic and reactive substances are proposed.

Keywords: Quantitative Structure-Property Relationships, *in silico* design, safety-by-design, physical hazards, energetic materials, reactive substances.

1. Introduction

Energetic materials are designed to be capable for releasing the chemical energy stored in their molecular structure upon external stimulations (like heat, shock or electrostatic discharges) in a short time[1]. Typical classes of energetic materials are propellants, explosives and pyrotechnic compositions that can be found in both military and civil applications, like propellants for aerospace, automotive airbags and entertainment fireworks. Reactive materials are substances which can cause fire or explosion by reacting by itself or in contact with other substances[2, 3]. They include different kinds of substances like organic peroxides, water-reactive or self-reactive substances.

Due to the hazards of these classes of substances, particular attention is needed in view of their safe storage and handling as they caused large numbers of major accidents among years[2, 4, 5]. For this reason, experimental tests have been developed to help assessing their risk profiles and risk assessments are required for their production, storage and use in regulations like, at international level, in the UN Recommendations for the Transport of Dangerous Goods (TDG)[6] and in the Globally Harmonized System of classification and labelling of chemicals (GHS)[7]. Based on these test results, substances are classified (e.g. according to the GHS regulation in Table 1) and specific labelling and safety measures are imposed to limit the occurrence and effects of accidents.

With decades of experience in this field, Ineris represents a first rank international expert in the assessment and characterization of fire and explosion risks. Its experimental facilities allow characterizations from the lab-scale to real scale explosion testing in a dedicated pyrotechnic platform. These tests are realized according to regulations and standards but are also adapted to investigate specific scenarios with tailor-made apparatus and protocols.

If a first tentative had been conducted in the 1990s with the works of Nefati[8] on the prediction of the impact sensitivity of explosive substances, the characterization of physical hazards at Ineris was up this time focused on the experimental approach. Taking advantage of the improvement of computer capacities and of theoretical methods and encouraged by the regulatory context of the European REACH regulation[9], new researches were engaged in 2007 in collaboration with Chimie ParisTech (Pr. Adamo) to investigate again the potential of chemoinformatics and theoretical chemistry to complement the experimental approach to predict the explosive properties of nitrocompounds. Based on the success of this work, this research activity has been continued and structured in a research roadmap with two objectives: 1) the development of prediction methods of physical hazards of substances and mixtures and 2) the theoretical study of hazardous chemical reactions. Among prediction methods, Quantitative Structure-Property Relationships (QSPR) are particularly targeted. This research has been supported by a series of national and international projects. In addition to a fruitful scientific production, this research produced tools and methods that are now used at Ineris in its support to the French Ministry in charge of Environment and in its expertise for industries as complementary tools to its experimental facilities.

This paper aims to propose a review of Ineris' contribution to the development and use of QSPR models in the field of substance and process safety. After an introduction to the principles and methods used for the development of QSPR models, an overview of the models developed at Ineris for the physical hazards of energetic and reactive materials will be provided, including the efforts engaged to make this approach applicable to mixtures. Then, discussion will be proposed on the use of QSPR models

taking into account some specificities for the case of physical hazards. At last, perspectives and challenges will be identified.

2. Methods

If Quantitative Structure-Activity Relationships (QSAR) are used from decades in the field of (eco)toxicology[10-12], applications to physical hazards were rare, with only few examples on explosive and flammability properties[13-16], and developments in this field are relatively recent[17]. When applied to physico-chemical properties (or hazards), such model may be more likely called QSPR, for Quantitative Structure-Property Relationships, since the target endpoint is a property instead of a biological activity.

The QSPR method consists in developing models by correlating an experimental property of chemical compounds with a series of molecular descriptors as summarized in Eq. 1, considering the assumption that compounds with similar molecular structures present similar properties.

$$Property = f(Descriptors) \quad (1)$$

The development of such model requires the availability of an experimental database that has a critical importance for the development and validation of models[18]. These data should be as large and as accurate as possible. Indeed, uncertainties and errors in the training data propagate in the model during its fitting and affect its final performances. Moreover, they should be obtained into homogeneous protocols because different protocols can give different and sometimes contradictory data that can limit the access to reliable models. Finally, the dataset also defines the detailed endpoint of the model (in terms of experimental protocol and conditions) and its applicability domain (in terms of properties and chemical space).

Such databases are not easily available in the field of physical hazards. For this reason, efforts were needed to consolidate the databases used for the development of models for these properties. Indeed, if some properties like the flash point or flammability limits are relatively largely provided for pure compounds in the DIPPR[19] or the ChemSafe[20] databases, consolidated datasets are more difficult to find for some other properties (like explosive properties) or specific substances (like self-reactive substances and mixtures). For this reason, the collection and analysis of available data in literature represents an important working step in the development of models. Dedicated experimental campaigns were even specifically conducted in some cases[21, 22].

Molecular descriptors encode the molecular structure of the chemical compounds. Thousands were developed among the decades[23, 24], calculated from simple elemental formula, from the 2-dimension or 3-dimension structures (after quantum chemical optimizations). Different types of descriptors can be found:

- Constitutional descriptors, like the occurrence and count of particular features in molecules (atoms, bonds, fragments, functional groups);
- Topological descriptors encoding molecular structures from molecular graphs and related to their size, shape or ramification rate;
- Geometric descriptors that characterize the 3D structure of molecules, like interatomic distances, angles, dihedral angles, molecular volume and surface areas;

- Quantum chemical descriptors obtained from quantum chemical calculations and related to electrostatic properties (e.g. atomic charges) or reactivity (e.g. bond dissociation energies).

The choice of the descriptors used to develop models is dependent on the target chemicals, on the property under investigation and on the level of complexity expected for the models. Indeed, simple models only based on descriptors that can be calculated from the SMILES codes or the count of functional groups on a 2D molecular structures are easier to use but can be less predictive and chemical comprehensive than models including quantum chemical descriptors presenting a chemical insight with the property under investigation. For this reason, several software has been used for the descriptors calculation, mainly Codessa[25] and Dragon[26], based on different types of input molecular structures, from SMILED codes to 3D structures optimized using the density functional theory (DFT) as implemented in Gaussian program[27]. Some external descriptors were also proposed, like Oxygen Balances[6, 28], widely used in the safety of explosives, or quantum chemical descriptors related to reactivity (like conceptual DFT descriptors[29]) or to specific functional groups like atomic charges in peroxy or nitro bonds. It should also be kept in mind that the interpretation of QSPR models in terms of chemical properties depends on the descriptors on which it is based. The chemical interpretation of some of them (notably topological indices) are not fully intuitive and limited to general aspects like the size, compacity or level of ramification of molecules, which can make the structure-property relationship, in the field of physical hazards, less straightforward than models focused on specific behaviors of functional groups related to explosive properties.

Different forms of models can be found depending on the method used to derive it[30, 31]: artificial neural networks (ANN), partial least squares (PLS), support vector machines (SVM), decision trees and multilinear or nonlinear regressions (MLR and NLR, respectively). If multi-linear regressions are very simple to finally use, some others like neural networks could need more implementation efforts to be used once developed. Multi-linear regressions can be also relatively easy to analyze in terms of chemical interpretation of the trends associated to the descriptors involved in the model, even if more complex methods like ANN can be more reliable to account for non-linear trends. Decision trees are also easy to use and interpret for qualitative predictions in classes of hazards, like “HIGH” vs “LOW”.

The descriptors included in the final model are selected to avoid at best any over-parameterization of the model that would lower its predictive power. Several data mining strategies can be used [32]. For instance, in the case of MLR models, stepwise methods, like the Best Multi Linear Regression technique (BMLR) implemented in Codessa[25], or genetic algorithm (GA), as in QSARINS[33, 34], can be used to evidence the set of descriptors representing the best compromise between the quality of fit of the model and the chemical meaning of the included descriptors. In general, only small datasets are often available in the field of physical hazards. For this reason, the selection of descriptors is of critical importance and models included few chemical comprehensive descriptors were in general preferred to highly parameterized models.

Because predictions of physical hazards can be used in regulatory context or for risk assessment purpose, particular attention was paid to the validation of the models. A series of internal and external validations are performed[35-38]. Leave-on-out and leave-many-out cross-validations are performed to evaluate the robustness of the models. Y-randomization methods [39] are used to ensure against chance correlation. The predictive power of models is then evaluated on an external validation set using a series of the statistical indices: R^2_{ext}, MAE_{ext}, $RMSE_{ext}$, Q^2_{F1}[40, 41], Q^2_{F2}[42], Q^2_{F3}[43], CCC[44], r^2_m and Δr^2_m[37].

When analyzing the performances of models for physical hazards, it should be kept in mind that only small datasets are in general available. So, validation sets are in general small too. This impacts not only the performances of models but also the precision of statistical indices themselves. For this reason, some models presenting the highest validation indices were not finally selected because the improvement of statistical performances was not large enough compared to models presenting a better interpretation or being easier to use.

At last, the applicability domain (AD) of the models, i.e. the domain in which predictions can be expected to be accurate, is systematically defined. The definition of AD can be done using different kinds of tools as reviewed by Jaworska et al.[45] or Eriksson et al.[46]. One of the simplest ways to proceed is to define the ranges of values of property and descriptors (for those included into the model) in the training set but other methods were also used like the leverage approach as implemented in QSARINS software[33, 34].

3. Development of QSPR models for the physical hazards of energetic and reactive materials

3.1. Energetic materials: models for nitro compounds

The first developments engaged in 2007 focused on the explosive properties of nitro compounds. A series of models were developed based on different datasets found in literature for different properties involved in the characterization of potentially explosive substances. Based on a small dataset of 22 nitroaromatic compounds[47], first studies targetted the heat of decomposition, a pre-selection criterion for the classification of explosive substances[6]. If this dataset was not large enough to achieve an external validation of models, it allowed initiating the setup of the method, notably by investigating the relevance of quantum chemical descriptors[48, 49] and, in particular, those issued from conceptual DFT[50]. From then, models were developed for different physico-chemical hazardous properties and for different families of compounds for which prediction models were lacking or for which improvement of existing prediction models was possible.

On the heat of decomposition (ΔH) of nitro compounds, the availability of a larger dataset[51] allowed the obtention of models with an external validation to evaluate its predictive power. The most accurate ones focused on nitrobenzene derivatives presenting no substituent in ortho position to the nitro group[52-54]. Indeed, theoretical studies on the decomposition of nitrobenzene[55] and o-nitrotoluene derivatives[56] demonstrated nitroaromatics with ortho substituent could present specific decomposition mechanisms. So, considering only non-ortho substituted compounds allowed focusing on compounds following, at best, a same subjacent reaction mechanism. The best models obtained based on this assumption consisted in the four-parameter multilinear regression presented in Eq. 2, based on quantum chemical descriptors calculated by Codessa[25] from molecular structures optimized at PBE0[57]/6-31+G(d,p) level.

$$-\Delta H = 0.8\ G - 3.8\ WPSA1 - 4255.1\ Qmax + 26.8\ RPCS - 251.2 \qquad (2)$$

with *G* the gravitational index, *WPSA1* the weighted positive surface area, *Qmax* the maximal partial charge and *RPCS* the relative positively charged surface area.

This model was fitted on a training set of 31 molecules with a mean absolute deviation of 12% and a R^2 of 0.90 after a descriptor selection performed using the BMLR technique as implemented into

Codessa software[25]. Its robustness was demonstrated by cross-validation with R^2_{cv}=0.86. The predictive capacities of the model, evaluated on the validation set of 11 molecules, were also good with an observed average error of 18% and R^2_{ext} = 0.84.

To favor simpler application, requiring no preliminary time-consuming quantum chemical calculations, a model only based on constitutional descriptors was also developed (Eq. 3)[54].

$$-\Delta H = -594.5 + 2381.6\, n_{db,rel} + 306.5\, n_{NO2} - 791.4\, n_{O,rel} + 83{,}4\, n_{conj} \quad (3)$$

where $n_{db,rel}$ and $n_{O,rel}$ are the relative numbers of double bonds and oxygen atoms, respectively, and n_{NO2} and n_{conj} are the numbers of nitro groups and conjugated bonds.

As illustrated in Figure 1, the performances of this model are close to the ones of quantum chemical one with R^2 = 0.84, R^2_{cv} = 0.79, R^2_{ext} = 0.84 and an average error of 18% in the validation set.

In addition to the performances of models, their chemical interpretation is also an important consideration. On this aspect, both models include descriptors related (directly or indirectly) to nitro groups (e.g. n_{NO2}) that are central in the decomposition reactions of nitro compounds[55, 56] and recognized as a group associated to explosive properties[6].

A series of models was also developed for the impact sensitivity ($h_{50\%}$) of nitro compounds from data extracted for the experimental compilation of Storm[58] (all obtained using a single experimental protocol by the same laboratory). This property characterizes the tendency of the material to react under the effect of an impact. Different QSPR models were developed for nitroaromatics, nitroaliphatics and nitramines, even if validated models were only found for the last two families of compounds, as shown in Table 2 for models including quantum chemical descriptors (calculated by Codessa[25] from molecular structures optimized at PBE0/6-31+G(d,p) level)[59]. The models for nitramines and nitroaliphatics (in Eqs. 4 and 5, respectively) present good performances in terms of goodness-of-fit (with R^2 = 0.92 and 0.93), robustness (with Q^2_{LOO} = 0.89 and 0.90) and predictive capacities (with R^2_{ext} = 0.88 in both cases). But, if correlation and robustness could appear as also satisfying for nitroaromatics (with R^2 = 0.92 and Q^2_{LOO} = 0.75), external validation revealed no predictive capacity with R^2_{ext} = 0.01.

$$\log h_{50\%} = -0.42 - 0.017\, OB + 0.06\, T^E_1 + 50.1\, N_{O,avg} + 27.6\, N_{N,avg} \quad (4)$$

$$\log h_{50\%} = -0.438 - 0.018\, OB + 4.07\, P + 28.5\, Q^2_{NO2,max} + 4.76\, N_{O,max} \quad (5)$$

where OB is the oxygen balance as defined in the TDG regulation[6], T^E_1 is the topological electronic index, $N_{O,avg}$ and $N_{N,avg}$ are the average nucleophilic reactivity indices for a O and a N atom, P is the polarity parameter, $Q_{NO2,max}$ is the maximum charge for a NO_2 group and $N_{O,max}$ is the maximum nucleophilic reactivity index for a O atom.

As for the heat of decomposition, simpler models were also developed for the impact sensitivity of nitramines[60] and nitroaliphatics[61] (in Eqs. 6 and 7, respectively).

$$\log h_{50\%} = 0.94 + 86.3\, n_{C=O}/M_w - 0.017\, OB + 0.14\, n_{C\text{-}O\text{-}C} - 0.21\, n_{C=O} \quad (6)$$

$$\log h_{50\%} = 1.94 - 2.53\, n_{N,rel} + 0.07\, n_{single} + 50.1\, n_{NO2} \quad (7)$$

where $n_{C=O}$, n_{C-O-C} and n_{NO2} are the numbers of C=O, C-O-C and NO_2 fragments, *OB* is the oxygen balance as defined in the TDG regulation[6], $n_{N,rel}$ is the relative number of nitrogen atoms, n_{single} is the number of single bonds.

The performances of these simpler models were also good with R^2_{ext} of 0.81 for instance for nitroaliphatics. For nitramines, the simpler model in Eq. 6 presented even similar results as the quantum chemical one with a R^2_{ext} of 0.90 (vs. 0.88 for Eq. 4) and a $RMSE_{ext}$ of 0.14 (vs. 0.16) estimated on an external validation set.

Among works engaged on nitrocompounds, the development of a model dedicated to the electric spark sensitivity (E_{ES}) of nitramines may also be noticed [62]. This property is defined as the sensitivity of the energetic materials when subjected to a high-voltage discharge from a capacitor. Indeed, initiation modes was less explored than impact. It depends strongly on experimental protocol (e.g. configuration of electrodes). Moreover, the micro-mechanisms involved during this initiation remain not fully elucidated[63, 64]. This model (detailed in Eq. 8) was well correlated with $R^2 = 0.90$ but, based on a small dataset of only 26 compounds [65], no external validation was possible.

$$E_{ES} = 29.6\, n_{single} + 63.3\, N_{C,max} + 168.4\, Q_{C,min} - 27.8\, V_{C,min} + 99.4 \quad (8)$$

where n_{single} is the relative number of single bonds and $N_{C,max}$, $Q_{C,min}$ and $V_{C,min}$ are the maximum nucleophilic reactivity index, the minimum charge and valence for a carbon atom.

3.2. Reactive materials: models for organic peroxides and self-reactive substances

Models on the hazardous properties of organic peroxides were developed within the PREDIMOL project (2010-2014) [21, 66], which aimed to evidence the potential of molecular-scale modeling for the prediction of physicochemical properties in the framework of the European REACH regulation[9]. Organic peroxides are defined by the presence of unstable O-O bonds in the molecule (as shown in Figure 2). They are thermally unstable, can easily decompose and may lead to explosion. For this reason, they are classified in a specific division in regulations like the UN Recommendations for the Transport of Dangerous Goods (TDG)[6] and the Globally Harmonized System of classification and labelling of chemicals (GHS)[7]. For this class of compounds, a lack of dataset relevant for the development of QSPR models was evidenced encouraging the setup of a dedicated experimental campaign.

From the experimental data obtained from this campaign, models were developed for the heat and temperature of decomposition (ΔH and T_{onset}) of organic peroxides[21]. For the 38 tested organic compounds, more than 300 descriptors were calculated by Codessa based on the structures optimized at PBE0/6-31+G(d,p) level. Using the BMLR technique, the models Eqs. 9 and 10 were then derived for the heat and temperature of decomposition respectively.

$$-\Delta H/C = 54\, {}^1\kappa - 990\, n_{OO} + 12934\, d_{OO} + 2631\, Q_{OO} - 19371 \quad (9)$$

$$T_{onset} = 144\, F^-_{OO} + 29\, n_{OO} - 20\, gap + 194 \quad (10)$$

where C is the concentration in organic peroxide, ${}^1\kappa$ is the order 1 Kier shape index, n_{OO} is the number of peroxide bonds, d_{OO} is the length of the peroxide bond, Q_{OO} and F^-_{OO} are the average Mulliken charges and local Fukui functions on the oxygens of the peroxide bond, respectively, and gap is the difference of the LUMO and HOMO energies.

These models were satisfactory from a statistical point of view with notably respective correlations R^2 of 0.97 and 0.84 and predictivity R^2_{ext} of 0.81 and 0.83 in their applicability domain. They are also chemically interpretable with descriptors chemically relevant, because related to the peroxy bonds, whose homolytic cleavage is critical in the decomposition mechanism of organic peroxides[67, 68].

Once again, models focused on simple descriptors were also searched and a relevant model was found for the heat of decomposition (in Eq. 11). Even if slightly less predictive than the quantum-chemical one in Eq. 9 with R^2_{ext} = 0.80 in its applicability domain, it is easier to use since it is only based on descriptors that can be manually determined from the 2D structures of the target organic peroxide.

$$-\Delta H/C = -663\, n_{OO} - 699\, n_{OOH} - 4.79\, OB + 11\, n_{single} - 2036 \quad (11)$$

where n_{OO} and n_{OOH} are the numbers of O-O and O-OH bonds, *OB* is the oxygen balance according to the TDG regulation[6] and n_{single} is the number of single bonds.

Beyond the models developed within the PREDIMOL project, these data were also used by other teams notably to develop QSPR models based on other methods and tools[69, 70].

Work on organic peroxides has also been continued, in the HAZPRED project[22], with aim to extend investigations to other properties of organic peroxides and in particular to the self-accelerated decomposition temperature (SADT) which is not only used as a classification criterion but also to determine the control temperature possibly required for safe transport[6]. On this property, few QSPR models were proposed in literature [71-73]. The obtained performances are already promising, despite some possible lacks in the used datasets in term of details on the tested samples and on the experimental protocols. Indeed, some commercial OP are mixtures, some are strongly diluted in solvent and different protocols can be used to determine the SADT. Nevertheless, they encourage further efforts to improve the data consolidation on organic peroxides.

In this project, self-reactive substances were also studied[22]. These materials are defined as thermally unstable substances liable to undergo a strongly exothermic decomposition without participation of oxygen (air). No QSPR model dedicated to this class of compounds was found in literature, due to the lack of enough consolidated experimental data. For this reason, data collection was engaged among literature. Moreover, an experimental campaign was conducted at BAM (Germany) on a series of samples provided by Bayer. These results were used in the first QSPR studies on these substances. At first, a partial least square (PLS) analysis was performed using SIMCA P+[74] on 45 substances based on more than 1900 molecular descriptors calculated by Dragon software[26] from SMILES codes. This preliminary study encouraged the development of models focusing on the heat of decomposition, that have been engaged using a GA-MLR methods as implemented in QSARINS software. The models issued from this last study will be soon published and demonstrate encouraging performances (considering the small quantity of available experimental data) to access first estimations of the heat of decomposition of self-reactive substances for instance for screening purpose to optimize experimental campaigns.

3.3. Towards mixtures

If the development of QSPR models in the recent years enlarged the availability of models to predict physico-chemical properties[75] and physical hazards[17, 76-78], this approach remains still mainly limited to pure compounds. Indeed, only few recent works are dedicated to mixtures[79], despite the fact that

the substances used in industrial processes are often mixtures. For this reason, efforts were engaged at Ineris on the use of QSPR models for the physical hazards of mixtures starting with the flash point (*FP*) of organic mixtures.

The flash point is the temperature at which the vapor above a flammable liquid ignites under the effect of a pilot flame[1] and is a key flammability property of liquids in the assessment of process safety issues and in various regulatory frameworks related to the transport, storage and use of chemicals[6, 7, 9, 80].

Numerous models exist to predict the flash point of pure compounds[81]. Besides, two models were developed in the PREDIMOL project specifically for amines and organic peroxides[82], even if the concept of flash point as to reflect the flammability hazard of organic peroxides is limited due to potential decomposition issues, in comparison with conventional flammable liquids.

In the case of mixtures, thermodynamic mixing rules can be used to estimate the flash point of mixtures based on the knowledge of those of its constituents. In 1982, Gmehling and Rasmussen[83] proposed to combine a vapor-liquid equilibrium model (UNIFAC[84]) and the Le Chatelier rule (to account for the flammability of vapor mixtures). More recently, this method was adapted by Liaw[85] and successfully applied to various types of mixtures including ternary[86], partially miscible[87] and aqueous mixtures[88], even if some limits were also evidenced, notably for chlorinated mixtures[89]. Moreover, these mixing rules require knowledge on data on their constituents, and in particular their individual flash points.

To solve this problem, a first approach tested by Ineris consisted in using QSPR models to provide the flash point of the pure compounds. For instance, the combination of QSPR models like the one of Carroll[90] with the mixing rule of Liaw demonstrated reliable estimations of flash points for binary mixtures[91], including miscible as partially miscible ones, and revealed also relevant for ternary mixtures[92].

In a second approach, the QSPR methodology for pure compounds was adapted to develop new QSPR models for the flash point of mixtures by introducing mixture descriptors in the models[93]. These mixture descriptors D_{mix} were defined using mixture formula combining the molecular descriptors d_i of the different constituents of the mixture and their respective molar fraction x_i in the mixture as presented in Eq. 12.

$$D_{mix} = f(x_i, d_i) \quad (12)$$

Based on 9 mixture formula and more than 250 molecular descriptors (calculated with Codessa), more than 2400 mixture descriptors were calculated. The used dataset was composed of 650 flash point data for 60 binary mixtures (in different concentrations), found in literature and selected to be obtained in homogeneous conditions using closed cup apparatus in accordance with regulatory requirements[6, 94]. A specific attention was paid to the partition of the dataset into the training and validation sets. Indeed, as pointed out by Muratov et al.[79], the presence of the same mixture in both sets (even in different concentrations) could lead to an over-estimation of the predictive power of the model. Using a “mixture-out” partition, no mixture of the validation set was also represented in the training set. Even if the obtained model[95], presented in Eq. 13, offers lower performances (with a MAE of 7.3°C in prediction) than the use of QSPR predictions to obtain the flash points of the pure compounds for inclusion in mixing rules (MAE = 2.9°C[91]), this approach could be useful when no mixing rule can be used, e.g. for other properties or in absence of relevant liquid-vapor equilibrium model.

$$FP(°C) = 20.3 + 28.6\ \Sigma(x_i\ T^E_i) + 24.6\ \Sigma(\sqrt{x_i}\ {}^3\chi) + 59.6\ \Sigma(\sqrt{x_i}\ V_{/XYZ}) + 315.0\ \Sigma(\sqrt{x_i}\ Q_{H,max}) - 107.6\ \Sigma(\sqrt{x_i}\ V_{H,min}) + 2.0\ (\Sigma(x_i\ \mu))^2 \qquad (13)$$

where T^E is the topographic electronic index (all bonds), ${}^3\chi$ is the Randic index (order 3), $Q_{H,max}$ and $V_{H,min}$ are the maximum partial charge and the minimum valence for a H atom, μ is the dipole moment and $V_{/XYZ}$ is the molecular volume/XYZ box.

Besides, this approach and the proposed mixture formula have been recently successfully used for other properties like lower flammability limits[96], auto-ignition temperatures[97] and even the mixture toxicities of antibiotics and pesticides[98].

4. Favoring extended and robust use of QSPR models for safety applications

In addition to its research for the development of QSPR models dedicated to physical hazards, Ineris engaged efforts on their diffusion and towards their robust use in regulatory context and for safety applications.

4.1. Robust application of QSPR model

The safety of reactive materials is strongly framed by regulations like the United Nations TDG and GHS Recommendations[6, 7] or the European REACH regulation[9]. In the perspective of use of QSPR predictions into regulatory context, OECD (Organization for Economic Co-operation and Development) proposed five validation principles[40] that guided the methodology of development of QSPR models for physical hazards at Ineris.

Principle 1. A defined endpoint. This definition includes the experimental protocols and conditions associated to the data used to develop the model. Even if not always possible, a single protocol should be used for sake of homogeneity. Concerning physical hazards, it may be checked because different standard protocols exist and can be adapted to investigate specific process situation. For instance, in the case of flash point, closed cup measurements are in general more conservative than open cup ones by about 10°C[1] and are in general recommended for the regulatory classification of flammable liquids[1]. For this reason, the models developed for the flash point[93, 95] were fitted only using closed cup data.

Principle 2. An unambiguous algorithm. This principle favors the correct application of the model in future uses. It includes the details of the algorithm of the model and the computational details related to the calculation of the descriptors, e.g. the quantum chemical level of theory. In the case of commercial tools (that implement the model or descriptors), the full algorithm can be complex and/or proprietary. In such case, it is important to provide not only the parameters used in the tool when used but also its version. Indeed, software are subject to updates to improve algorithms, to increase their reliability, enlarge their field of application or improve their computer performances or ergonomic aspects. So, different versions of a same software can provide different answers. For this reason, the version of all software must be also clarified.

Principle 3. A defined applicability domain. The definition of the applicability domain can be not limited to statistical considerations in terms of property and descriptor values but also in terms of subjacent mechanisms. For instance, the domain of applicability of model in Eq. 2 is limited to the

prediction of heats of decomposition for non-ortho substituted nitroaromatics due to the possible deviation in the decomposition mechanisms in presence of ortho substituents[53].

Principle 4. Appropriate measures of the correlation, robustness and predictive power of the model. In this principle, OECD recommends the external validation of models based on an external test set, in addition to internal validation methods like cross-validation and Y-randomization. This external validation may nevertheless be done with caution because the availability of a relevant test set is sometimes difficult. Indeed, in the field of physical hazards, the available datasets are relatively small. Moreover, in the case of mixtures, the definition of the test set should include at best mixtures that do not include constituents present in the training set, as recommended by Muratov et al.[79].

Principle 5. An interpretation of subjacent mechanisms, if possible. To fulfill this principle, models including chemical comprehensive descriptors were favored. For instance, the models developed for decomposition properties (heat, temperature of decomposition, impact sensitivity) include descriptors related to molecular features involved in the decomposition mechanism of the target compounds (NO_2 for the nitro compounds and -O-O- for organic peroxides).

It is worth noting that these principles do not give acceptance criteria of models ensuring that all predictions issued from them should be considered as reliable and accepted. Besides, there is no unique measure of model acceptability and it can depend on the context in which the model is applied. They intend to highlight the robustness of the model development and to give the information needed to justify a correct and relevant application of models in view of regulatory acceptance[40]. Additional guidance may be followed to ensure the good use and report of QSPR predictions, as those proposed by the European Chemicals Agency (ECHA)[99-101].

At first, a QSPR model is based on the identification of correlations among a defined set of data. For this reason, it can be considered as predictive only in a defined applicability domain in terms of chemical diversity and in terms of property domain. Predictions out of its training domain (by extrapolation) can't be considered as reliable. Moreover, the target of the model and the final prediction should be relevant for the expected use of the predicted data (especially in regulatory context). So, to achieve robust predictions, the following steps should be successfully fulfilled.

Step 1: Choice of the model.

At first, the model should be relevantly chosen. Of course, it might at best fit the validation principles proposed by OECD. But it must also be selected in terms of target endpoint and chemical diversity. For instance, within a regulatory context, the model should target properties as determined according to specific protocols. So, they should be trained on data obtained according to these requirements. This is notably the reason why the model developed for the flash point of mixtures (in Eq. 13) was trained for experimental data obtained according to the closed cup ASTM D58 standard which is among the recommended protocols in regulations[6, 94].

The scope of the model must also be checked in term of target substances. Some models, even large, can be not applicable to particular compounds, depending on the composition of its training set. Others are focused on specific ones like the model presented in Eq. 2 dedicated to heat of decomposition of nitroaromatic compounds without ortho substituent. In this last model, this limitation is important since ortho substituents may affect the decomposition reactions of

nitroaromatics. So, users should check if the structural features (e.g. functional groups or meaningful fragments) of the target chemical are represented in the training set and can even enforce it if the subjacent mechanism is known to be the same for the target chemical than for those of the training set.

Step 2: Correct application of the model.

Then, the model must be applied according to the exact procedure defined by its developers. This concerns not only the algorithm but also the way its descriptors are calculated. In the case of quantum chemical descriptors, computational details, like the basis set and the method (e.g. functional), must be followed.

Step 3: Check of the applicability domain.

A refined analysis of the applicability domain should be done. At first, it may be checked if the descriptor values for the studied compound belong to the applicability domain as defined by the authors. Unfortunately, only recent models present such applicability domain definition[76]. However, such analysis can be done afterwards based on the descriptor values of the compounds in the training set. For instance, it can be checked if the descriptors are calculated within the ranges of descriptor values within the training set. It should be similarly checked if the predicted value also belongs to the range of property values of the training set.

Step 4 : Analysis of structural analogues.

The presence of close structural analogues in the training set of the model can be considered as increasing the reliability of the prediction. Moreover, the prediction for these compounds can give a better idea of the level of accuracy of the prediction. The more similar is an analogue and the more similar should be the level of accuracy of its prediction.

This can be also complemented by the search of additional structural analogues out of the training and validation sets that could be more similar to the target compounds and which reliability of prediction could be even more relevant than analogues of the training set to estimate the reliability of the prediction. Nevertheless, it should be kept in mind that the model may target a specific protocol. So, the experimental data found on the structural analogue must be issued from a relevant protocol. For instance, in the case of the flash point, to evaluate a prediction issued from the model in Eq. 13, the experimental data on the analogues should be obtained by a closed cup apparatus, as those of the training set of the model.

Step 5: Analysis of the adequacy of the prediction.

This step consists in checking if the obtain prediction can be used. Beyond the relevance of the target endpoint of the models, a prediction can reveal not adequate for a decision making depending on its value, its level of reliability and the nature of the decision to be taken.

For instance, only the more accurate predictions should be considered when close to a regulatory threshold whereas larger uncertainty can be enough when the prediction is largely conservative compared to the threshold. For example, heat of decomposition is one of the pre-selection criteria for explosive substances in the UN Manual of tests and criteria[6]. If a substance presents a heat of

decomposition lower than 500 J/g, it should not be considered as eligible for classification as explosive substance. So, for this property a predicted value largely higher than 500 J/g could be easily considered adequate to engage the full experimental procedure to evaluate to possible classification of the substance as explosive (class 1) whereas a prediction slightly lower than 500 J/g should be of very high accuracy to be enough to exclude a possible classification in class 1.

A precautionary principle should be recommended when considering this adequacy. Prediction with the highest uncertainty and the most severe impact in case of errors should be considered in the most conservative way, i.e. avoiding possible underestimation of the hazard, by applying safety margins or by affecting the more severe hazard level.

Finally, to demonstrate that models and predictions satisfy these recommendations, specific documentations can be used. As illustrated in Figure 3, the QSAR model reporting format (QMRF) summarizes the compliance of the model with the OECD principles whereas the QSAR prediction reporting format (QPRF) summarizes all information that demonstrate the relevance of predictions for the expected use[99].

At this point, it may also be kept in mind that QSPR predictions, even when adequacy is good, can always be complemented by other additional experimental or theoretical arguments to enforce decision making. For instance, when several applicable models are available, the results obtained from the different relevant models can be analyzed. This is particularly relevant when the different models are independents in terms of algorithms and training set. This concept is notably exploited in so-called consensus models [102, 103] that combine different models to give a final prediction. Unfortunately, for most of physical hazards, even when several models exist, they often are issued from the same database.

Beyond the application for the classification of hazardous substances, such approach can also be used to provide unknown safety data in process safety analysis. One interesting case concerns the flammability of liquid mixtures. When the flash point of a mixture is unknown, a commonly used assumption, thought as conservative, consists in taking the flash point of the most flammable constituent as the one of the mixtures. This assumption can be good for mixtures presenting an ideal behavior but fails for mixtures presenting a minimum flash point behavior, as for the n-octane/ethanol mixture, as shown in Figure 4. In such case, the flash point of the mixture is lower than the one of the pure compounds meaning a higher flammability. On the contrary, the predictive method combining mixing rule and QSPR models succeeds to account for this specific behavior and provides good estimations of the flash point (with an observed mean error of 1.9°C).

4.2. Diffusion and training activities

Even now, the number of models available for the prediction of physical hazards keeps limited in comparison with the large number of models existing for (eco)toxicological properties. So, efforts have been engaged in the diffusion of models to favor their use by industrials and regulators. Indeed, up to now, none of the 154 models listed in the JRC QSAR database[104] (in March 2020) concerns physical hazards. If many free and commercial tools include QSAR/QSPR models to predict (eco)toxicological and even some thermophysical properties like boiling point or octanol-water partition coefficient[77], physical hazards are rarely represented. The only exceptions concern some flammable properties of gas and liquids (flammability limits, flash point and self-ignition temperature).

For this reason, Ineris initiated efforts for the diffusion of some of its models in such tools. This process has been initiated with the OECD/ECHA QSAR Toolbox [105]. This free software aims to increase the regulatory acceptance of QSAR/QSPR methods to make them more accessible, transparent, and lowly demanding in terms of infrastructure costs[106]. It notably includes QSAR and QSPR models implemented after acceptance by an expert comity on their scientific validity and the technical feasibility of their implementation.

Considering this last point, only the simpler models of Ineris were proposed for integration in the QSAR Toolbox. The first proposed model was the model based on constitutional descriptors dedicated to the prediction of the impact sensitivity of nitroaliphatic compounds[61] presented earlier in Eq. 7. In addition to its simplicity (for implementation and use), this model was chosen because its target endpoint is of regulatory interest and because it complies with the five OECD validation principles as described below:

Principle 1. A defined endpoint. This model is dedicated to the impact sensitivity of nitroaliphatic compounds which is one of the most important property to classify explosive substances according to the CLP[80] and TDG[6] regulations. All the data were extracted from the compilation of data proposed by Storm et al.[58] that gathered experimental results issued from a same protocol, by drop weight impact test.

Principle 2. An unambiguous algorithm. The final model is the multilinear regression presented in Eq. 7. The descriptors can be extracted from the simple 2D molecular structure and were selected using the stepwise Best Multi Linear Regression algorithm in Codessa software[25].

Principle 3. A defined applicability domain. This model is applicable to nitroaliphatics presenting impact sensitivities between 6 and 300 cm. In terms of descriptor domain, their range of applicability are the following:

- relative number of nitrogen atoms: 0.118 to 0.250;
- number of single bonds: 11 to 44;
- number of nitro groups: 2 to 12.

Principle 4. Appropriate measures of the correlation, robustness and predictive power of the model. This model was validated using a series of internal and external validations that demonstrated good performances in terms of:

- goodness-of-fit: R^2=0.88 and RMSE = 0.17;
- robustness by cross-validation: Q^2_{LOO}=0.85, Q^2_{10CV}=0.84 and Q^2_{5CV}=0.85;
- chance correlation by Y-randomization: R^2_{YS}=0.09 and SD_{YS}=0.07;
- predictive power by external validation: R^2_{ext}=0.81, RMSE = 0.22, Q^2_{F1}=Q^2_{F2}=0.81, Q^2_{F3}= 0.83 and CCC=0.93.

Principle 5. An interpretation of subjacent mechanisms, if possible. Concerning chemical interpretation, the number of nitro groups is particularly to be noticed. At first, nitro groups are among the functional groups identified as associated to explosive properties[6]. Secondly, the primary cleavage of the C-NO_2 bond is known as critical in the decomposition mechanism of nitroaliphatic compounds[107].

After examination by the expert comity, this model was accepted and implemented into version 3.3 of the QSAR Toolbox in December 2014. Two other simple models were proposed recently by Ineris: for the impact sensitivity of nitramines (Eq. 6) and for the heat of decomposition of organic peroxides (Eq. 11). They also comply with the OCDE principles and are simple enough to be implemented in the Toolbox. They were also evaluated as relevant for integration and their availability could be expected in a future version of the QSAR Toolbox.

These three models were also proposed for integration in the QSAR DataBank (QsarDB)[108, 109]. Managed by the Institute of Chemistry of Tartu University (Estonia), this platform lists more than 200 models for both (eco)toxicological and physico-chemical properties. Each model is provided with the associated QMRF file and the datasets used for its development when available. It even allows predictions from some of them. In this case, a prediction module calculates the descriptors, the predicted property value and analyses of their compliance with the AD of the model. It also provides a list of similar compounds found in the dataset used to develop and validate the model to evaluate the reliability of the predictions on structural analogues. Finally, it allows the edition of a first draft of QPRF file that can be completed by the users.

The three models have been integrated in the platform[110-112]. The details on the models are provided as the full databases used to derive the models. It also allows search for data in the databases and identification of structural analogues. For the model for nitroaliphatics[61, 110], it is even possible to perform predictions from the prediction module.

To favor the diffusion of good practices in the development and use of QSPR models, Ineris also engaged in the sensibilization of industrial users and regulators on methods based on structural similarity (QSAR/QSPR, read across). Training sessions are proposed to potential users and evaluators of predictions for both toxicological and physico-chemical properties.

4.3. Towards *in silico* design applications

The development of new substances, devices or processes is often costly and time-expensive. This is particularly the case when hazardous substances are involved. Indeed, in such cases, risk assessment should be conducted before being placed on the market based on sometimes important experimental test campaigns. Besides, these tests are needed not only in the perspective of proposing a product on the market but also for the risk assessment associated to their production, storage and use. In general, physical hazards are only tested for the solutions identified as the most relevant for the target application after various R&D steps focused on the expected functionalities and economic considerations. But if these solutions revealed unsafe, safety barriers should be added. In the worse cases, the product must be completely canceled, ruining months of R&D efforts.

The availability of predictive method gives the opportunity to anticipate physical hazards earlier in the development of new substances in a safer-by-design approach. Indeed, the use of QSPR models do not require any experiments and estimations of properties can be obtained even before any synthesis. For this reason, they can give first estimation of physical hazards to eliminate as earlier as possible candidates, presenting too severe hazard potentials. They also represent interesting screening tools to select among a series of potential candidates the ones with possibly the lowest safety constraints even before synthesis and experimental characterizations.

It can be for instance used for the identification of safer formulations. Predictive methods have notably been used to define the diluent concentration required to lower the flammability of dimethylamine (DMA) aqueous solutions[113]. 40% DMA solution presents a very low flash point (below 0°C), corresponding to a very high flammability level. As a consequence, important safety barriers are required for its storage. A way to reduce this flammability is the dilution in water to reduce the concentration in DMA. Predictions highlighted that increasing water concentration can help to increase its flash point and that a 5% solution would present a flash point of about 32°C which would change the classification of the solution from Flammable liquid of category 2 to category 3 (in the European CLP regulation[80]) and reduce the constraints related to its storage and use. These theoretical results encouraged further experimental test on this 5% DMA solution that confirmed the relevance of this new concentration for safer storage.

5. Conclusions and Perspectives

Chemoinformatics is a fruitful field of research for different application domains. The recent developments conducted on energetic materials and for physical hazards demonstrates the potential of such approaches as a tool to anticipate safety issues when using potentially hazardous substances, devices and processes. After about a decade of research, the methods and models developed by Ineris are fully part of the tools used by the institute in its expertise for industries and regulators, complementary to experimental testing.

However, the QSPR approach presents some limitations that must be considered:

- QSPR models are correlations methods. So, they are only applicable for compounds similar to their training set. Moreover, their performances are limited by the accuracy of the experimental data used to develop them;
- Almost only the most recent models (developed in the last two decades) fulfilled the OECD validation principles for use in regulatory context, in particular concerning the external validation and the definition of the applicability domain;
- QSPR models are based on molecular descriptors. For these reasons, they do not account for polymorphism and the influence of solid-state structures;
- QSPR models do not account for the temperature or pressure dependency of the predicted property. Models are in general developed for a property in specific experimental conditions (e.g. the flash point at standard pressure conditions);
- If some models are dedicated to mixtures, impurity effects are not considered;

Notably for these reasons, some scientific challenges remain:

- Despite the increase of works published in this field, most of the properties associated to physical hazards remain to be investigated in structure-property studies. For instance, no model exists for oxidizing properties or for the corrosivity for metals. Some of them could besides represent methodological challenges (in particular for solid state properties);
- Existing models remain mainly limited to the most common families of compounds (like nitroaromatics or nitramines among explosive compounds). But research (on energetic materials, notably) investigates a larger diversity of compounds, like high nitrogen content energetic materials[114] or energetic salts[115]. So, the search of models for emerging energetic materials could be waited;

- The development of models for mixtures is still at its beginning and its application of some properties of energetic materials, in particular in the case of solid (with non homogeneous mixing like in liquids or gas), could be particularly challenging;
- At last, improvement of models could also be expected when considering the small size of the databases on which they were trained. The consolidation of extended reliable database could represent a real driver towards the improvement of models and the extension of their application fields. Large database could offer better fitting of the models and open better opportunity for deep learning methods more adapted to large datasets.

It may be noticed that chemoinformatics studies conducted at Ineris for the prediction of physical hazard are complemented by quantum chemical tools in process safety evaluation. At first, quantum chemical descriptors are calculated based on density functional theory calculations, as illustrated in this paper. Moreover, the detailed reaction mechanisms of hazardous reactions are investigated theoretically. These mechanistic studies favor a better understanding of these hazardous phenomena at a molecular level. For instance, the peroxidation of ethers was deeply analyzed to evidence the main reaction steps leading to the accumulation of peroxide species in long term storage of biofuel formulations, that decrease the fuel performances and can even cause accidents due to the reactivity of organic peroxides. Once the main reactivity scheme evidenced for the diethyl ether[116], application to other ethers[117] was performed to evaluate their relative potential for peroxide accumulation. Then, inhibition reactions were also studied to identify the best peroxidation inhibitor for a given ether among the ones commonly used in industrial formulations[118]. Reactions between substances are also investigated, as in the case of chemical incompatibility. For instance, pure ammonium nitrate is stable in normal conditions but can become hazardous is contact of incompatible substance. This incompatibility caused several major accidents. By comparing the detailed decomposition mechanisms of pure ammonium nitrate [119] with its reaction in presence of various additives and contaminants presenting incompatible (like Dichloroisocyanuric Acid [120]) or compatible behavior (like calcium carbonate [121]), new insight in the compatibility or incompatibility process were figured out. These theoretical studied can be used to complement experimental knowledge on the risks associated to NA formulations as used in the fertilizer industry.

At last, it should be reminded that these theoretical tools are complementary to the experimental facilities available at Ineris for the characterization of physical hazards of energetic materials and reactive substances and of hazardous reactions. Similar expertise is also available on (eco)toxicology at Ineris which allows the possibility to propose extended investigations on the complete risk profiles of substances and industrial processes for the human health and environment issued from accidental or chronic hazardous phenomena, using a versatile toolbox of experimental and modeling methods.

Acknowledgments

These research activities have been mainly supported by the French Ministry in charge of Ecology (Ministère de la Transition Ecologique et Solidaire). The authors thank all students and researchers who directly or indirectly contributed to the works cited in this review. The authors particularly thank Pr. Carlo Adamo (Chimie ParisTech) for his long-term and fruitful scientific collaboration.

References

[1] V. Babrauskas, *Ignition Handbook*, Fire Sciences Publishers, **2003**.

[2] T. Yoshida, Y. Wada, N. Foster, *Safety of Reactive Chemicals and Pyrotechnics, Vol. 5*, Elsevier, Amsterdam, **1995**.

[3] *Bretherick's Handbook of Reactive Chemical Hazards (Seventh Edition)*, Academic Press, Oxford, **2006**.

[4] E. K. Mihailidou, D. Antoniadis, M. J. Assael, *Int. Rev. Chem. Eng.* **2012**, *4*, 529-540.

[5] J. Sales, F. Mushtaq, M. D. Christou, R. Nomen, *Process Saf. Environ. Protect.* **2007**, *85*, 117-124.

[6] United Nations, *Recommandations on the Transport of Dangerous Goods: Manual of Tests and Criteria, 7th Revised Edition*, New York and Geneva, **2019**.

[7] United Nations, *Globally harmonized system of classification and labelling of chemicals (GHS), 7th Revised Edition*, New York and Geneva, **2017**.

[8] H. Nefati, J.-M. Cense, J.-J. Legendre, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 804-810.

[9] European Commission, *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)*, **2006**.

[10] *ECETOC Technical Report No. 74: QSARs in the Assessment of the Environmental Fate and Effects of Chemicals,* European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC),, Brussels, **1998**.

[11] S. P. Bradbury, *Toxicol. Lett.* **1995**, *79*, 229-237.

[12] T. I. Netzeva, M. Pavan, A. P. Worth, *QSAR Comb. Sci* **2008**, *27*, 77-90.

[13] A. R. Katritzky, R. Petrukhin, R. Jain, M. Karelson, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1521-1530.

[14] Y. S. Kim, S. K. Lee, J. H. Kim, J. S. Kim, K. T. No, *J. Chem. Soc. Perkin Trans. 2* **2002**, 2087-2092.

[15] S. R. Saraf, W. J. Rogers, M. S. Mannan, *J. Hazard. Mater.* **2003**, *98*, 15-29.

[16] D. V. Sukhachev, T. S. Pivina, F. S. Volk, *Propel. Explos. Pyrotech.* **1994**, *19*, 159-164.

[17] Z. Jiao, H. U. Escobar-Hernandez, T. Parker, Q. Wang, *Process Saf. Environ. Protect.* **2019**, *129*, 280-290.

[18] J. C. Dearden, M. T. D. Cronin, K. L. E. Kaiser, *SAR QSAR Environ. Res.* **2009**, *20*, 241-266.

[19] *Project 801, Evaluated Process Design Data, Public Release Documentation, Design Institute for Physical Properties (DIPPR)*, American Institute of Chemical Engineers (AIChE), **2006**.

[20] *CHEMSAFE-Database for Recommended Safety Characteristics, 2013-2015*, BAM, PTB and DECHEMA, **2020**, http://dechema.de/en/chemsafe.html.

[21] V. Prana, P. Rotureau, G. Fayet, D. André, S. Hub, P. Vicot, L. Rao, C. Adamo, *J. Hazard. Mater.* **2014**, *276*, 216-224.

[22] G. Fayet, K.-D. Wehrstedt, A. Knorr, P. Rotureau, *Chem. Eng. Trans.* **2019**, *77*, 337-342.

[23] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, **2000**.

[24] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, **2000**.

[25] Codessa, University of Florida, **2018**, http://www.semichem.com/codessa/default.php.

[26] Kode srl, Dragon (Software for Molecular Descriptor Calculation), 7.0, Pisa, Italy, **2016**, https://chm.kode-solutions.net.

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

Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian16, Gaussian Inc., Wallington CT, **2016**, http://www.gaussian.com.
[28] E. S. Shanley, G. A. Melhem, *Process Saf. Prog.* **1995**, *14*, 29-31.
[29] P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* **2003**, *103*, 1793-1874.
[30] J. Gasteiger, T. Engel, *Chemoinformatics - A textbook*, Wiley, Weinheim, **2003**.
[31] I. H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco, **2005**.
[32] M. Shahlaei, *Chem. Rev.* **2013**, *113*, 8093-8103.
[33] QSARINS, 2.2.3, Varese, Italy, **2018**, http://www.qsar.it/.
[34] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, *J. Comput. Chem.* **2013**, *34*, 2121-2132.
[35] A. Tropsha, *Mol. Inform.* **2010**, *29*, 476-488.
[36] P. Gramatica, *QSAR Comb. Sci* **2007**, *26*, 694-701.
[37] K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das, H. Kabir, *J. Chem. Inf. Model.* **2012**, *52*, 396-408.
[38] N. Chirico, P. Gramatica, *J. Chem. Inf. Model.* **2012**, *52*, 2044-2058.
[39] C. Rücker, G. Rücker, M. Meringer, *J. Chem. Inf. Model.* **2007**, *47*, 2345-2357.
[40] *Guidance Document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models,* Organisation for Economic Co-operation and Development (OECD), **2007**.
[41] A. Tropsha, P. Gramatica, K. V. Gombar, *QSAR Comb. Sci* **2003**, *22*, 69-77.
[42] G. Schüürmann, R. U. Ebert, J. Chen, B. Wang, R. Kühne, *J. Chem. Inf. Model.* **2008**, *48*, 2140-2145.
[43] V. Consonni, D. Ballabio, R. Todeschini, *J. Chem. Inf. Model.* **2009**, *49*, 1669-1678.
[44] N. Chirico, P. Gramatica, *J. Chem. Inf. Model.* **2011**, *51*, 2320-2335.
[45] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, *Altern. Lab. Anim.* **2005**, *33*, 445-459.
[46] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramatica, *Environ. Health Persp.* **2003**, *111*, 1361-1375.
[47] Y. S. Duh, C. Lee, C. C. Hsu, D. R. Hwang, C. S. Kao, *J. Hazard. Mater.* **1997**, *53*, 183-194.
[48] G. Fayet, P. Rotureau, L. Joubert, C. Adamo, *J. Hazard. Mater.* **2009**, *171*, 845-850.
[49] G. Fayet, P. Rotureau, L. Joubert, C. Adamo, *J. Mol. Model.* **2010**, *16*, 805-812.
[50] G. Fayet, L. Joubert, P. Rotureau, C. Adamo, *Chem. Phys. Lett.* **2009**, *467*, 407-411.
[51] T. Ando, Y. Fujimoto, S. Morisaki, *J. Hazard. Mater.* **1991**, *28*, 251-280.
[52] G. Fayet, A. Del Rio, P. Rotureau, L. Joubert, C. Adamo, *Mol. Inform.* **2011**, *30*, 623-634.
[53] G. Fayet, P. Rotureau, L. Joubert, C. Adamo, *J. Mol. Model.* **2011**, *17*, 2443-2453.
[54] G. Fayet, P. Rotureau, C. Adamo, *J. Loss Prevent. Proc. Ind.* **2013**, *26*, 1100-1105.
[55] G. Fayet, L. Joubert, P. Rotureau, C. Adamo, *J. Phys. Chem. A* **2008**, *112*, 4054-4059.
[56] G. Fayet, L. Joubert, P. Rotureau, C. Adamo, *J. Phys. Chem. A* **2009**, *113*, 13621-13627.
[57] C. Adamo, V. Barone, *J. Chem.* **1999**, *110*, 6158-6170.
[58] C. P. Storm, J. R. Stine, *Sensitivity Relationships in Energetic Materials,* in *Chemistry and Physics of Energetic Materials* (Ed.: S. N. Bulusu), Kluwer Academic Publishers, Netherlands, **1990**, pp. 605-630.
[59] G. Fayet, P. Rotureau, V. Prana, C. Adamo, *Process Saf. Prog.* **2012**, *31*, 291-303.
[60] G. Fayet, P. Rotureau, *J. Loss Prevent. Proc. Ind.* **2014**, *30*, 1-8.
[61] V. Prana, G. Fayet, P. Rotureau, C. Adamo, *J. Hazard. Mater.* **2012**, *235-236*, 169-177.
[62] G. Fayet, P. Rotureau, L. Joubert, C. Adamo, *Predicting explosibility properties of chemicals from Quantitative Structure-Property Relationships*, in *Proceedings of the 43rd Loss Prevention Symposium and 5th Global Congress on Process Safety* (Ed.: AICHE), **2009**, pp. 15-33.
[63] S. Zeman, *A study of chemical micro-mechanisms of initiation of organic polynitro compounds,* in *Energetic Materials. Part 2: Detonation, Combustion* (Eds.: P. Politzer, J. S. Murray), Elsevier, Amsterdam, **2003**, pp. 25-52.
[64] F. Hosoya, K. Shiino, K. Itabashi, *Propel. Explos. Pyrotech.* **1991**, *16*, 119-122.
[65] M. H. Keshavarz, *J. Hazard. Mater.* **2008**, *153*, 201-206.

[66] V. Prana, P. Rotureau, D. André, G. Fayet, C. Adamo, *Mol. Inform.* **2017**, *36*, 1700024.
[67] R. Benassi, U. Folli, S. Sbardellati, F. Taddei, *J. Comput. Chem.* **1993**, *14*, 379-391.
[68] R. Benassi, F. Taddei, *Tetrahedron* **1994**, *50*, 4795-4810.
[69] N. Zohari, M. H. Keshavarz, Z. Dalaei, *J. Therm. Anal. Calori.* **2016**, *125*, 887-896.
[70] Y. M. Dai, L. L. Niu, J. Q. Zou, D. Y. Liu, H. Liu, *J. Central South Univ.* **2018**, *25*, 1535-1544.
[71] Y. Gao, Y. Xue, Z. G. Lu, Z. Wang, Q. Chen, N. Shi, F. Sun, *Process Saf. Environ. Protect.* **2015**, *94*, 322-328.
[72] B. Wang, H. Yi, K. Xu, Q. Wang, *J. Therm. Anal. Calori.* **2017**, *128*, 399-406.
[73] Y. Pan, Y. Zhang, J. Jiang, L. Ding, *J. Loss Prevent. Proc. Ind.* **2014**, *31*, 41-49.
[74] SIMCA P+, Version 12.0, Umetrics AB, Umeå, Sweden, **2009**, https://umetrics.com/products/simca.
[75] A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn, D. A. Dobchev, *Chem. Rev.* **2010**, *110*, 5714-5789.
[76] C. Nieto-Draghi, G. Fayet, B. Creton, X. Rozanska, P. Rotureau, J.-C. De Hemptinne, P. Ungerer, B. Rousseau, C. Adamo, *Chem. Rev.* **2015**, *115*, 13093-13164.
[77] J. C. Dearden, P. Rotureau, G. Fayet, *SAR QSAR Environ. Res.* **2013**, *24*, 545-584.
[78] F. A. Quintero, S. J. Patel, F. Munoz, M. S. Mannan, *Ind. Eng. Chem. Res.* **2012**, *51*, 16101-16115.
[79] E. N. Muratov, E. V. Varlamova, A. G. Artemenko, P. G. Polishchuk, V. E. Kuz'min, *Mol. Inform.* **2012**, *31*, 202-221.
[80] European Commission, *Regulation (EC) N°1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) N° 1907/2006*, **2008**.
[81] J. R. Rowley, D. K. Freeman, R. L. Rowley, J. L. Oscarson, N. F. Giles, W. V. Wilding, *Int. J. Thermophys.* **2010**, *31*, 875-887.
[82] G. Fayet, P. Rotureau, V. Prana, C. Adamo, *Chem. Eng. Trans.* **2013**, *31*.
[83] J. Gmehling, P. Rasmussen, *Ind. Eng. Chem. Fund.* **1982**, *21*, 186-188.
[84] A. Fredenslund, R. L. Jones, J. M. Prausnitz, *AIChE J.* **1975**, *21*, 1086-1099.
[85] H.-J. Liaw, Y.-H. Lee, C.-L. Tang, H.-H. Hsu, J.-H. Liu, *J. Loss Prevent. Proc. Ind.* **2002**, *15*, 429-438.
[86] H.-J. Liaw, C.-L. Tang, J.-S. Lai, *Combust. Flame* **2004**, *138*, 308-319.
[87] H.-J. Liaw, W.-H. Lu, V. Gerbaud, C.-C. Chen, *J. Hazard. Mater.* **2008**, *153*, 1165-1175.
[88] H.-J. Liaw, Y.-Y. Chiu, *J. Hazard. Mater.* **2003**, *101*, 83-106.
[89] G. Fayet, B. Tribouilloy, P. Rotureau, *Process Saf. Prog.*, e12127.
[90] F. A. Carroll, C.-Y. Lin, F. H. Quina, *Ind. Eng. Chem. Res.* **2011**, *50*, 4796-4800.
[91] T. Gaudin, P. Rotureau, G. Fayet, *Fire Saf. J.* **2014**, *74*, 61-70.
[92] G. Fayet, P. Rotureau, *Biofuel Bioprod. Bior.* **2016**, *10*, 738-752.
[93] T. Gaudin, P. Rotureau, G. Fayet, *Ind. Eng. Chem. Res.* **2015**, *54*, 6596-6604.
[94] European Commission, *Regulation (EC) N° 761/2009 of 23 July 2009 amending, for the purpose of its adaptation to technical progress, Regulation (EC) N° 440/2008 laying down test methods pursuant to Regulation (EC) N° 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)*, **2009**.
[95] G. Fayet, P. Rotureau, *Mol. Inform.* **2019**, *38*, 1800122.
[96] Y. Pan, X. Ji, L. Ding, J. Jiang, *Molecules* **2019**, *24*, 748.
[97] Y. Jin, J. Jiang, Y. Pan, L. Ni, *J. Therm. Anal. Calori.* **2020**, *140*, 397-409.
[98] L.-T. Qin, Y.-H. Chen, X. Zhang, L.-Y. Mo, H.-H. Zeng, Y.-P. Liang, *Chemosphere* **2018**, *198*, 122-129.
[99] *Practical guide - How to use and report (Q)SARs,* Version 3.1, European Chemicals Agency (ECHA), **2016**.
[100] *Guidance Document on information requirements and chemical safety assessment, Chapter R.4: Evaluation of available information,* Version 1.1, European Chemicals Agency (ECHA), **2011**.

[101] *Guidance Document on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals,* European Chemicals Agency (ECHA), **2008**.
[102] I. Mitra, A. Saha, K. Roy, *J. Mol. Model.* **2012**, *18*, 1819-1840.
[103] P. Gramatica, E. Giani, E. Papa, *J. Mol. Graph. Model.* **2007**, *25*, 755-766.
[104] European Commission, Joint Research Centre (JRC), *JRC QSAR Model Database*, **2020**, http://data.europa.eu/89h/e4ef8d13-d743-4524-a6eb-80e18b58cba4
[105] OECD/ECHA, *QSAR Toolbox, version 4.4.1*, **2020**, http://www.qsartoolbox.org/.
[106] T. W. Schultz, R. Diderich, C. D. Kuseva, O. G. Mekenyan, *The OECD QSAR Toolbox Starts Its Second Decade,* in *Computational Toxicology. Methods in Molecular Biology, Vol. 1800* (Ed.: O. Nicolotti), Humana Press, New York, **2018**, pp. 55-77.
[107] G. Fayet, P. Rotureau, B. Minisini, *Phys. Chem. Chem. Phys.* **2014**, *16*, 6614-6622.
[108] V. Ruusmann, S. Sild, U. Maran, *J. Cheminformatics* **2015**, *7*, 32.
[109] QSAR DataBase, Tartu, Estonia, **2020**, www.qsardb.org.
[110] G. Piir, *QDB archive #209*, **2019**, http://dx.doi.org/10.15152/QDB.209.
[111] G. Piir, *QDB archive #231*, **2020**, http://dx.doi.org/10.15152/QDB.231.
[112] G. Piir, *QDB archive #230*, **2020**, http://dx.doi.org/10.15152/QDB.230.
[113] G. Fayet, P. Rotureau, B. Tribouilloy, *Chem. Eng. Trans.* **2019**, *77*, 781-786.
[114] T. M. Klapötke, *High Energy Density Materials*, Springer-Verlag, **2007**.
[115] H. Gao, J. n. M. Shreeve, *Chem. Rev.* **2011**, *111*, 7377-7436.
[116] S. Di Tommaso, P. Rotureau, O. Crescenzi, C. Adamo, *Phys. Chem. Chem. Phys.* **2011**, *13*, 14636-14645.
[117] S. Di Tommaso, P. Rotureau, C. Adamo, *J. Phys. Chem. A* **2012**, *116*, 9010-9019.
[118] S. Di Tommaso, P. Rotureau, W. Benaissa, P. Gruez, C. Adamo, *Energy Fuels* **2014**, *28*, 2821-2829.
[119] S. Cagnina, P. Rotureau, G. Fayet, C. Adamo, *Phys. Chem. Chem. Phys.* **2013**, *15*, 10849-10858.
[120] S. Cagnina, P. Rotureau, G. Fayet, C. Adamo, *Ind. Eng. Chem. Res.* **2014**, *53*, 13920-13927.
[121] E. Menicacci, P. Rotureau, G. Fayet, C. Adamo, *ACS Omega* **2020**, *5*, 5034-5040.
[122] H.-J. Liaw, V. Gerbaud, Y.-H. Li, *Fluid Phase Equil.* **2011**, *300*, 70-82.

Table 1 – Physical hazard classes in the GHS regulation[7].

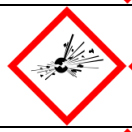
1	Explosives		10	Pyrophoric solids	
2	Flammable gases		11	Self-heating substances and mixtures	
3	Aerosols		12	Substances and mixtures which, in contact with water, emit flammable gases	
4	Oxidizing gases		13	Oxidizing liquids	
5	Gases under pressure		14	Oxidizing solids	
6	Flammable liquids		15	Organic peroxides	
7	Flammable solids		16	Corrosive to metals	
8	Self-reactive substances and mixtures		17	Desensitized explosives	
9	Pyrophoric liquids				

Table 2 – Performances of the quantum chemical based models developed for the impact sensitivities of nitrocompounds[59].

Target Compounds	R^2	Q^2_{LOO}	R^2_{ext}	$RMSE_{ext}$
Nitramines	0.92	0.89	0.88	0.16
Nitroaliphatics	0.93	0.90	0.88	0.19
Nitroaromatics	0.81	0.75	0.01	3.50

Figure 1 – Experimental vs. Predicted heats of decomposition (in kJ/mol) of nitroaromatics using quantum chemical and simple QSPR models (Eqs. 2 [53] and 3 [54], respectively)

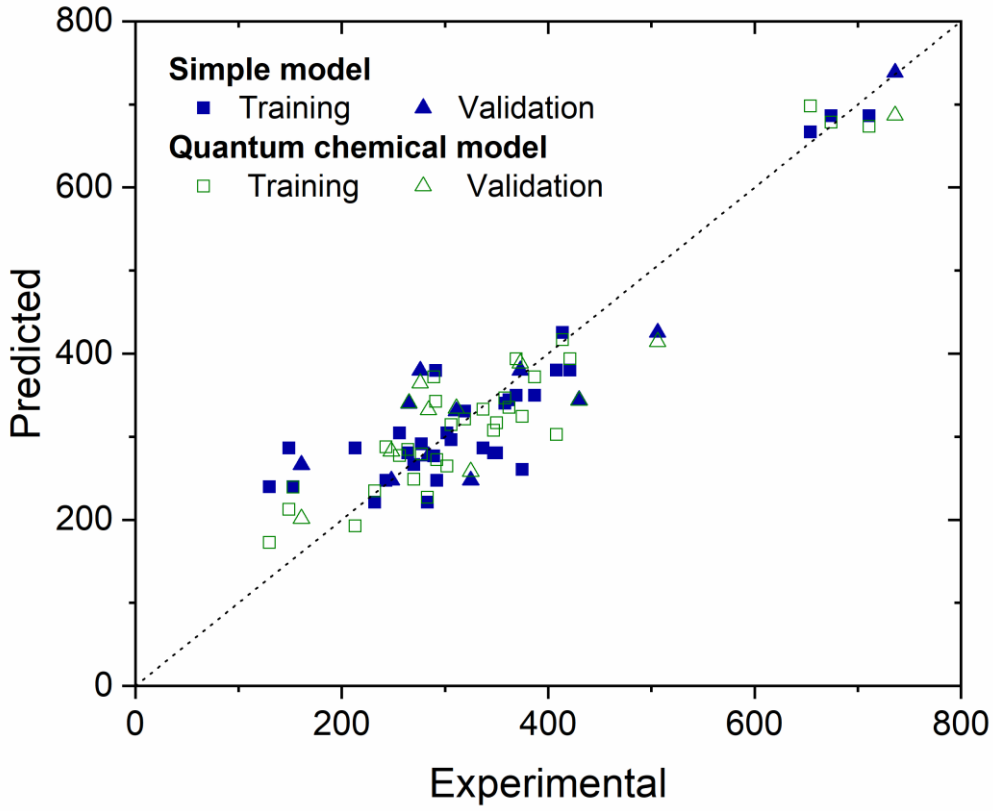
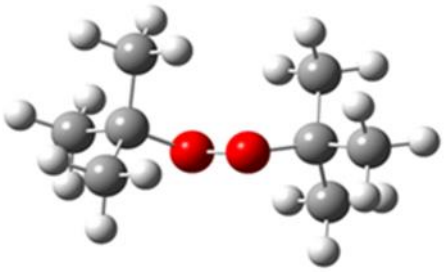
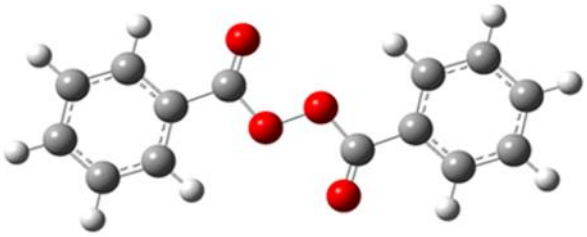


Figure 2 – Examples of organic peroxides



di-tert-butyl peroxide
(DTBP)



dibenzoyl peroxide
(DBPO)

Figure 3 – Illustration of the respective scopes of the QMRF and QPRF files

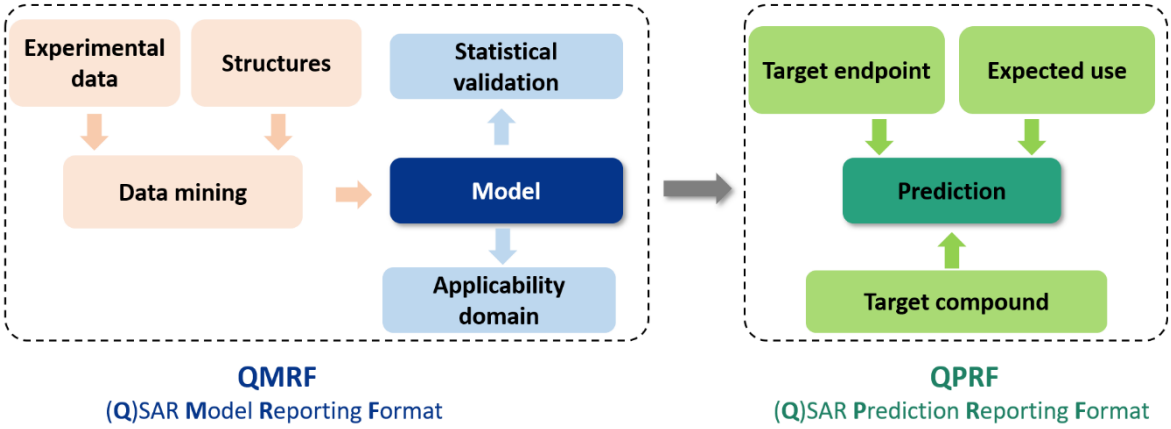
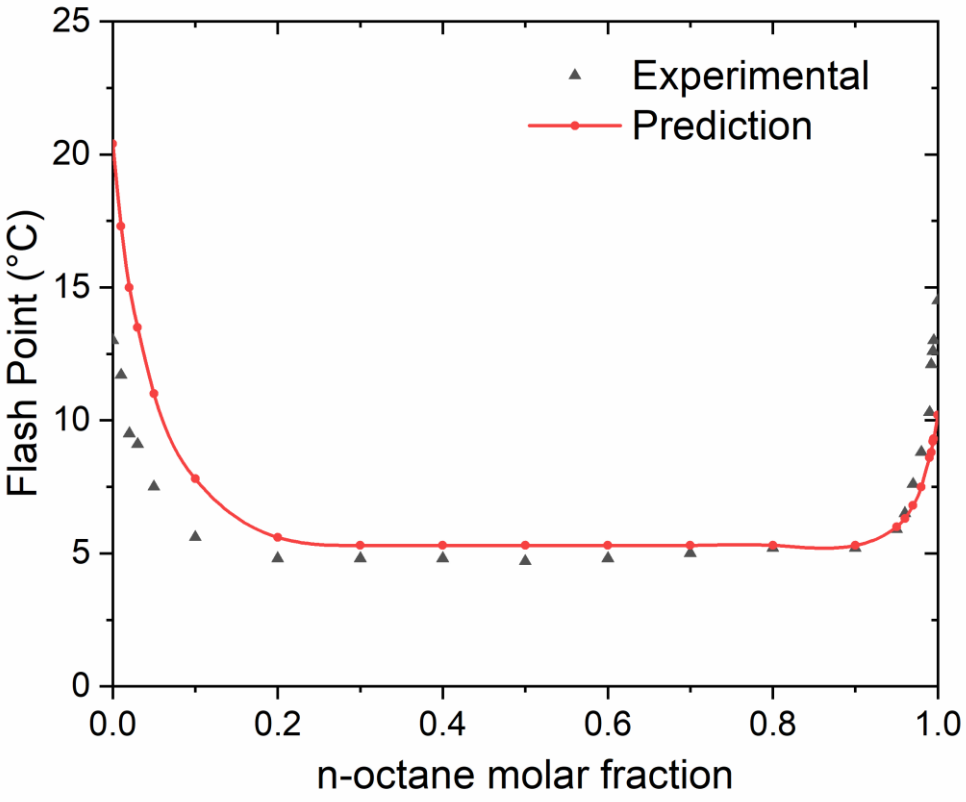



Figure 4 – Flash point of n-octane/ethanol mixtures measured (from Liaw et al. [122]) and predicted (by the full predictive method from Gaudin et al. [91])