CHEMICAL ENGINEERING TRANSACTIONS

VOL. 31, 2013

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at: www.aidic.it/cet

Guest Editors: Eddy De Rademaeker, Bruno Fabiano, Simberto Senni Buratti
Copyright © 2013, AIDIC Servizi S.r.l.,
**ISBN** 978-88-95608-22-8; **ISSN** 1974-9791

DOI: 10.3303/CET1331141

# Quantitative Structure-Property Relationships for Thermal Stability and Explosive Properties of Chemicals

Nadia Baati[a], Annik Nanchen[b], Francis Stoessel[b], Thierry Meyer[a]

[a]Ecole Polytechnique Fédérale de Lausanne, Institute of Chemical Sciences and Engineering, Group of Chemical and Physical Safety, Station 6, 1015 Lausanne, Switzerland.
[b]Swissi Process Safety, WRO 1055.5, 4052 Basel, Switzerland.
thierry.meyer@epfl.ch

Handling and storing chemicals in the industrial world require a conscientious hazards assessment and the implementation of appropriate measures to ensure safety for the workers, the company, the environment and the society. Reliable evaluations and predictions based solely on the molecular structure would represent a valuable tool in the preliminary hazard assessment process as it would reduce the necessary time and resources for extensive testing. This work presents how Quantitative Structure-Property Relationships (QSPR) of various hazardous chemical properties (i.e. thermal decomposition enthalpy or Minimal Ignition Energy) can be built to meet these needs. Results are to be illustrated with some examples. Two sets of chemicals were studied and the corresponding experimental results were correlated to 90 molecular descriptors. The models were built following a regression analysis. The best multi-linear regressions presenting 6 parameters or less with high determination and cross-validation coefficients are withheld as predictive models.

## 1. Introduction

Most of the chemicals encountered in the industry are in a metastable state and might decompose, releasing significant amounts of energy and byproducts. Thus, a very particular attention is required in order to properly tackle the incurred hazards and implement the appropriate safety measures. Therefore, a deep knowledge and understanding of the chemical reactions' characteristics and the involved compounds are crucial to properly assess the related risks and hazards to design a safe process.

Thermal stability can be defined by the energy and the kinetics of the decomposition reaction compounds can undergo. This can be determined experimentally by various thermal analysis methods. Using an adequate analysis, one can estimate the risks faced by integrating these reactions into a process and thus begin at an early process development stage, to think about the measures to implement in order to make it safer. Besides ensuring process safety through protective and preventive measures, these considerations can have deeper implications: if the incurred risks are too important, one could imagine a different design or process that would be inherently safer. It should be designed to avoid facing the threats rather than protect from their consequences. For this to be possible, process engineers need reactions and compounds characteristics at a very early development stage.

In this work two hazardous behaviours of chemicals will be considered: thermal decompositions and explosions. Thermal decompositions occur above their triggering temperature. As they are often exothermic, they induce a consequent temperature rise accelerating the reaction, leading to a possible thermal runaway. Heat accumulation could result in high temperature and pressure rises which may generate explosions. Some key characteristics of these reactions, if known, would help avoiding these hazardous situations. For instance, the process temperature should be lower than the temperature at which the reaction is triggered (onset temperature $T_o$) or the heat exchange system designed to manage the released reaction heat ($\Delta H_r$) and heat flux.

Please cite this article as: Baati N., Nanchen A., Stoessel F. and Meyer T., 2013, Quantitative structure-property relationships for thermal stability and explosive properties of chemicals, Chemical Engineering Transactions, 31, 841-846 DOI: 10.3303/CET1331141

In this paper chemical explosions resulting from rapid and violent oxidation reactions are considered, not the physical explosions due to overpressures within confined environments. Chemical explosions can be triggered by any ignition source.

This work aims at predicting results from Differential Scanning Calorimetry (DSC), which reveals the onset temperature $T_o$, the enthalpy change $\Delta H_r$ related to the decomposition reaction and gives strong indication on the decomposition kinetics. Minimal Ignition Energy (MIE) corresponds to the energy of an electric spark sufficient to induce a dust explosion and is determined by explosion tube tests.

The predictive models are produced as Quantitative Structure-Property Relationships (QSPR), linking up the experimental values to numerical molecular descriptors of the examined compounds. The employed molecular descriptors encode for several aspects of the compound’s structure such as constitution, shape or electronic properties. Correlations can be found by multi-linear regressions or non-linear approaches.

## 2. Methods

### 2.1 Quantitative Structure-Property Relationships (QSPR)

Several methods of predicting chemicals properties have been recorded in literature. Some rely on associating chemical property value to other physical properties of the compounds (Valenzuela et al., 2011) and some other to their molecular structure.

When basing predictive models on molecular structure, the historical strategy is derived from Benson’s group contribution. The American Society for Testing and Materials (ASTM) launched in 1974 a software tool to predict thermochemical properties and evaluate hazards related to substances from their structure based on Benson’s method (Shanley and Melhem, 1995). This technique has been applied until the mid-2000’s (Albahri, 2003; Nomen, 2005). Even though it performs well, this technique is limited to molecules that can be described by the group-additivity system and is unable to distinguish isomers.

Nowadays, QSPR are more often encountered since their efficiency has been proven in several publications. The QSPR methodology was first developed for biochemical studies as Quantitative Structure Activity Relationships (QSAR) enabling researchers to match the complex molecules structure and their biochemical activity (e.g. toxicities (Harder et al., 2003). Since then, QSPR have been applied for predicting several properties related to chemical’s reactivity or hazards: auto-ignition temperatures AIT (Pan et al., 2009), maximum overpressure $P_{max}$ and explosion constants $K_{st}$ for explosive dusts (Reyes et al., 2011), impact sensitivity (Xu et al., 2012) among others. Saraf (2003) and Fayet (2009) proposed models to predict onset temperature $T_o$ or reaction enthalpy $\Delta H_r$ by QSPR models. Later, Fayet (2011) also applied these techniques to explosibility properties of chemicals.

The first step in QSPR routine is to gather the experimental values of the property to be modelled. The second step consists into producing a numerous molecular descriptors for each of the studied compounds. The molecular descriptors can be of various kinds: constitutional (e.g. numbers of atoms and bonds), geometric (e.g. volume, solvent-accessible surface areas), topological (that depend on atoms connectivity), electrostatic (e.g. partial charges) or “quantum-chemistry derived” (e.g. molecular orbitals energies, dipole moments components, etc.) (Karelson, 2000). To generate these descriptors, variable software packages are available. They derive a three dimensional molecular structure and calculate the descriptors based on the Density Functional Theory (DFT). With Karelson’s molecular descriptor handbook (Karelson, 2000), a generator software was made available and was used for this project.

Finally, multi-linear correlations between those descriptors and the values of the chemical property considered are to be found. For that purpose, one usually applies classical linear algebra combined with statistical principles. The sample set is to be divided between a training and validation set. The training set is used to find the individual molecular descriptors coefficients to fit the property are. The validation set serves to assess whether or not the models perform successfully in predicting values. This random splitting process was repeated before each modelling, thus there are several training and validations sets comprising different samples.

Stepwise regressions are recommended as they allow selecting descriptors successively while verifying which contribute the most to improving the model. The models are here produced using MATLAB’s (Mathworks, 2010) stepwise function that allows building multi-variate predictive models.

Another approach is to develop an artificial neural network considering the molecular descriptors as inputs and the property values as targets. Even though this is a very efficient strategy to achieve high correlations, it should be combined to a selection process in order to screen among the numerous descriptors. Indeed, non-linear methods can adjust coefficients for all parameters however they do not eliminate the less relevant ones (Pan et al., 2009).

In both cases, the models are then evaluated on the precision of fitting the training data with determination coefficient ($R^2$) and the precision of predicting the validation data with cross-validation coefficient ($R_{cv}^2$).

### 2.2 Experiments and data extraction

Differential Scanning Calorimetry (DSC) is a widely used thermoanalytical technique. A sample and a reference are placed in a furnace and submitted to a temperature profile. The difference in heat flows from the sample and the reference are recorded against the temperature and are graphically represented as thermograms.

The interpretation of a DSC curve is a key step of the analysis as it allows correlating the observed features of the thermogram to the thermal events. Within the scope of this work, the most interesting events to study are chemical reactions, especially decompositions. They are characterized by important heat releases revealed with significant deviations of the curve from the baseline forming notable peaks. The temperature for which this deviation starts corresponds to the onset temperature $T_o$ and the area under the curve is directly proportional to the reaction enthalpy $\Delta H_r$.

These two key parameters have been modelled by the previously mentioned studies (Saraf et al., 2003; Fayet et al., 2009). However, the overall shape of the peak also reveals the kinetics of the reaction, indicating if the decomposition occurs slowly after the onset (broad peak) or rather violently (narrow peak). Thus, in this study, in addition to the onset temperature and reaction enthalpy corresponding to the main peak, the curve is entirely described with an ensemble of properties: each one of the three main peaks observed on every thermogram is assimilated to a Gaussian peak and characterized by its position, height and width.

In order to reconstitute the thermograms, the estimated values are subjected to the reverse processing and the thermograms represented as a sum of Gaussian-like curves.

The Minimal Ignition Energy (MIE) is a measure revealing how sensitive an explosive dust is to an electrical spark ignition and is key data for explosion risk assessment (Lerena and Suter, 2010). The MIE values are obtained from tube explosion tests conducted in modified Hartmann tubes. A dust sample is dispersed within an explosion vessel with pressurized air, and two electrodes connected to a circuit produce an electrical spark of known energy. If the spark induces observable flame propagation, the spark's energy is reduced until no flame propagation is observed for ten consecutive tests. The highest energy for which no flame propagation occurs is recorded. However, the sample weight is known to influence the ignition energy. Therefore, a range of dust sample weights are tested and the lowest energy is considered as the MIE.

For modelling the explosive dusts sensitivity, only the MIE value will be considered in this paper, thus no particular processing is required on experimental data before moving on to the modelling phase.

## 3. Results and discussion

### 3.1 Thermal stability

DSC experiments have been conducted on 19 nitroaromatic compounds, 16 and 3 samples respectively in training and validation sets (in order to respect the 80/20 proportion). The experimental conditions have been standardized to avoid any perturbing effects on the measures and the investigated temperature range is between 30°C and 500°C.

All the studied samples showed reactivity on this temperature range with important main exothermic peaks. Out of the 19 samples 16 were initially on a solid state and presented an endothermic peak caused by the substance fusion during the experiments. Besides, some thermograms showed another exothermic peak directly preceding the main peak, revealing complex reactive mechanisms (i.e. consecutive decomposition reactions).

All the graphical data have been then processed as explained previously in order to extract the data of interest, namely the three main peaks' position, height and width. For that purpose each peak is fitted by a Gaussian curve and its characteristics are withheld to describe the peak's shape. These properties are then correlated to the molecular descriptors and a multi-linear equation is to be determined for each modelled property.

Different models have been produced and selecting the best models was realized not only according to the determination and cross-validation coefficients but also on the number of parameters included. Indeed, the more parameters are integrated to the model, the higher the determination coefficient is. However, the models are then more complex and their physical significance is decreased. Mathematically, with a high dimension system, it would be possible to fit any sort of data with very fine adjustments; as the goal here is to achieve the best predictive, yet comprehensible, models rather than the highest possible fitting to the training data. The models were only retained if they achieve accurate fittings with a limited number of parameters.

For instance, Eq. 1 and 2 correspond to the models produced respectively for predicting the thermograms main peak height in terms of heat flow ($q_{max,1}$ [W/g]) and position in terms of temperature ($T_{max,1}$ [°C]).

$$q_{max,1} = - 9.7\ NH + 302.1\ RNSB - 0.1\ DPSA - 119.6\ V_{mid} - 62.9 \quad (1)$$

$$T_{max,1} = - 27.5\ SIC0 - 1.4\ S_M + 2.3\ PPSA1 - 1167.3\ FPSA1 - 547.4\ HNC + 1151.3 \quad (2)$$

Table 1 presents an explanation for the parameters symbols. The estimated values for the various properties are calculated from these models and the included descriptors and compared to the experimental values for evaluating the determination ($R^2$) and cross-validation ($R_{cv}^2$) coefficients. Figures 1 and 2 show the graphical representations of the previous examples.

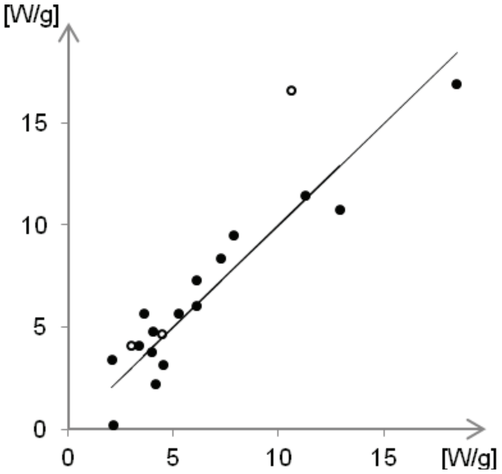

*Figure 1: Graphical view of calculated values $q_{max,1}$ vs experimental values – line shows perfect match of simulations and experiments-circles represent validation set- both axes are in [W/g]*

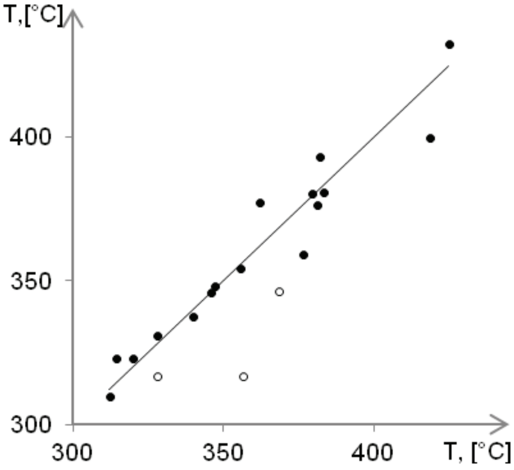

*Figure 2: Graphical view of calculated values $T_{max,1}$ vs experimental values – line shows perfect match of simulations and experiments- circles represent validation set-both axes are in [°C]*

For these two examples, the evaluation results were rather satisfactory as the $q_{max,1}$ model showed a $R^2$=0.96 and $R_{cv}^2$=0.83 and the $T_{max,1}$ model showed $R^2$=0.93 and $R_{cv}^2$= 0.84. All the properties of interest mentioned previously have been processed similarly and the results ranged between $0.71 \le R^2 \le 0.98$ and $0.67 < R_{cv}^2 < 0.96$ with a maximum of 6 descriptors integrated in a model.

When reconstructing the thermograms from the calculated values, the applied procedure achieved very interesting results with correlations up to $R^2$=0.96. For instance, the graphical comparison between measured and simulated thermograms of 4-nitroaniline is presented in Figure 3.

Overall, this technique proved to be efficient for predicting DSC results by rendering entire thermograms. The basic QSPR principles are applied, only the required data pre-processing is slightly more fastidious. Nevertheless, this enhancement on the current technique allows a complete estimation of the chemicals' thermal behavior instead of a characterization based on two values only.

### 3.2 Explosive sensitivity

Considering the successful application of QSPR methodology to predicting DSC experiments, it has been decided to apply it to MIE, addressing in the meanwhile the lack in predictive models for that property, which is, as previously mentioned crucial for explosion risk assessment. 50 chemicals have been tested on Hartmann modified test tubes. Their MIE were recorded and an ensemble of 80 descriptors was calculated for each compound. As the selected compounds were from various chemical classes, their MIE values were rather heterogeneous ranging from 0.015 mJ to 10 J. This very broad values span and unevenly dispersed data biased the modelling process. Therefore, the selected data set have been divided into three different sub-sets according to their MIE values, and different models were produced for each sub-set.

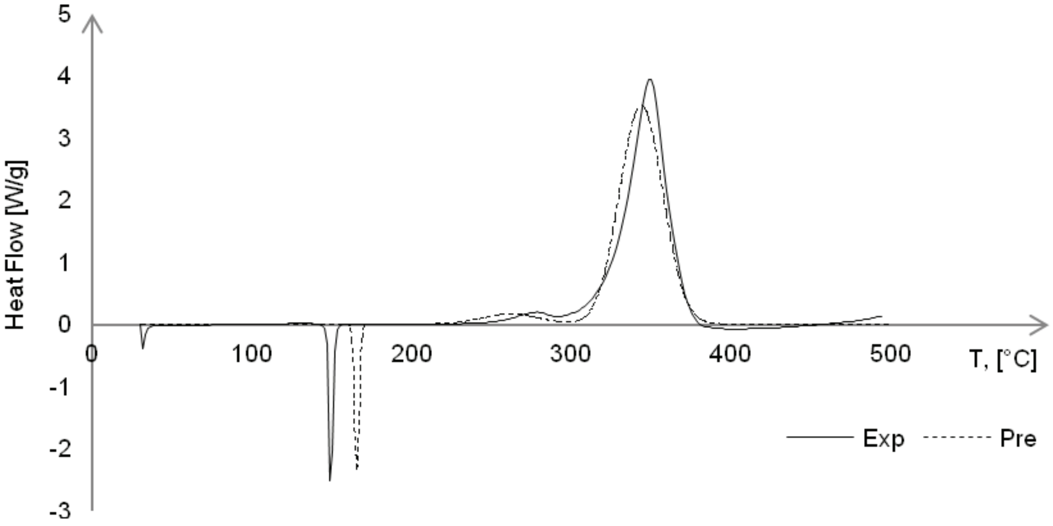

*Figure 3: Graphical comparison of 4-nitroaniline experimental (Exp) and predicted (Pre) thermograms*

To illustrate these results, Figure 4 below represents the comparison of experimental and calculated value of MIE between MIE = 30 mJ and MIE = 1000 mJ, which corresponds to a sub-set of 15 samples. These values were estimated with the following model and it presented $R^2$=0.98 and $R_{cv}^2$=0.97.

$$MIE = 114.1\ NH + 1772.8\ RNSB + 164.7\ {}^{3}\chi - 142.6\ \kappa^{3} - 6.6\ PPSA2 - 25702\ FNSA3 - 2171 \quad (3)$$

This model is a very interesting preliminary result. Indeed, the obtained correlation coefficients are high and the number of parameters low as desired. Nevertheless, only a restrained amount of data (15 samples) is taken into account here and this leads to questioning the reliability of the model. Even though the samples are separated into a training and a validation sets in order to increase the generalization abilities of the model, it is recommended that statistical models like these are derived from larger datasets to ensure higher robustness and better performances; otherwise it is not guaranteed that when applied to chemicals external to the studied set the model's predictive power would be as high as noted here.

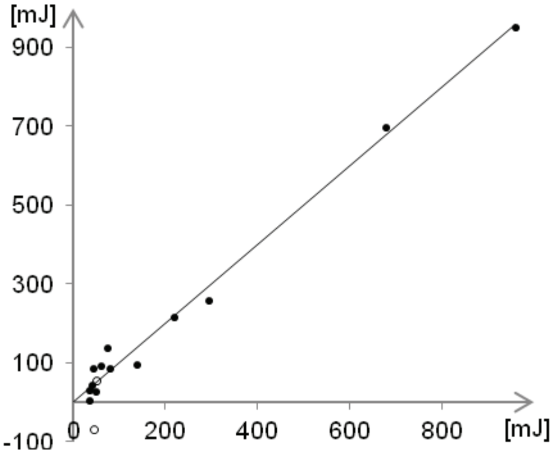

*Figure 4: Graphical view of calculated vs experimental MIE values – line shows perfect match of simulations and experiments- circles represent validation set- both axes are in [mJ]*

## **4.** Conclusions

For this work, QSPR models for characteristic properties of thermal stability and explosive sensitivity were developed. The properties were measured experimentally and the data of interest were extracted to be correlated to molecular descriptors. The relationships were built as multi-linear regressions that properly fit the experimental results. The technique was not solely applied to decomposition reactions onset temperature and enthalpy but also to the entire DSC thermograms in order to preserve the graphical information concerning the reaction kinetics. MIE values were also estimated with highly predictive model.

Nevertheless, this investigation reveals some challenges to face. Indeed, for predictive models to be as helpful as possible their reliability is crucial. Therefore, the training and validation sets should enclose more experimental data representing more diverse chemical classes and broader spans of measures.

*Table 1: Explanation of descriptors involved in Equations 1 to 3.*

| Symbol | Name | Explanation |
|---|---|---|
| NH | Number of H atoms | |
| RNSB | Relative Number of Single Bonds | |
| $^3\chi$ | Kier and Hall Index 3 | Topological descriptors extracted from molecular graph, intended to encode for different aspect of the molecule's shape |
| $\kappa^3$ | 3D Kier shape Index 3 | |
| SIC0 | Structural Information Content | |
| $S_M$ | Molecular Surface area | |
| HNC | Highest Partial Negative Charge | |
| $V_{mid}$ | Mid-point potential | Electrostatic potential at mid-point of C-NO2 bonds |
| PPSA1 | Partial Positive Surface Area 1 | Sum of partial positively charged solvent accessible areas |
| PPSA2 | Partial Positive Surface Area 2 | Product of Total charge and PPSA1 |
| FPSA1 | Fractional PPSA1 | Sum of partial positively charged solvent accessible areas relatively to total molecular solvent accessible area |
| FNSA3 | Fractional atomic charge weighted partial negative area | Product of Total charge and PNSA1 relatively to total molecular solvent accessible area |
| DPSA1 | Difference between partial positively and negatively charged surface areas | |

## References

Albahri, T., 2003. Flammability characteristics of pure hydrocarbons. Chemical Engineering Science, 58(16), pp.3629–3641.

Fayet, G., Rotureau, P., Joubert, L., Adamo, C., 2009. QSPR modelling of thermal stability of nitroaromatic compounds: DFT vs AM1 calculated descriptors. Journal of Molecular Modelling, 16, pp.805–812.

Fayet, G., Rotureau, P., Joubert, L., Adamo, C., 2010. Predicting Explosibility Properties of Chemicals from Quantitative Structure-Property Relationships. Process Safety Progress, 29(4), pp. 359-371.

Harder, A., Escher, B.I., Schwarzenbach, R.P., 2003. Applicability and limitation of QSAR for the toxicity of electrophilic chemicals. Environmental Science & Technology, 37(21), pp.4955–61.

Karelson, M., 2000. Molecular Descriptors in QSAR/QSPR, volume 1, Wiley-Intescience.

Lerena P., Suter, G., 2010. Tools to assess the explosion risks in the chemical, pharmaceutical and food industry. Chemical Engineering Transactions, 19, 243-248.

The MathWorks, 2010. MATLAB 7.10, Natick, MA.

Nomen, R., Barta, M., Sempere, J., Serra, E., Sales, J., Romero, X., 2005. The estimation of reaction enthalpy for complex molecules using Benson groups: Comparison of different strategies. Journal Of Thermal Analysis, 82, pp.581–584.

Pan, Y. Jiang, J., Wang, R., Cao, H., Cui, Y., 2009. Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine. Journal of Hazardous Materials, 164(2-3), pp.1242–9.

Reyes, O.J., Patel, S.J., Mannan, M. S., 2011. Quantitative Structure Property Relationship Studies for predicting dust explosibility characteristics (Kst, Pmax) of organic chemical dusts. Industrial and Engineering Chemistry Research, pp.2373–2379.

Saraf, S.R., Rogers, W.J., Mannan, M. S., 2003. Prediction of reactive hazards based on molecular structure. Journal of Hazardous Materials, 98(1-3), pp.15–29.

Shanley, E.S., Melhem, A., 1995. A review of ASTM evaluation criteria. Science, 8(5), pp.261–264.

Valenzuela, E.M. Vázquez-Román, R., Patel, S. J., Mannan, M. S., 2011. Prediction models for the flash point of pure components. Journal of Loss Prevention in the Process Industries, 24(6), pp.753–757.

Xu, J., Zhu, L., Fang, D., Wang, L., Xiao, S., Liu, L., Xu, W., 2012. QSPR studies of impact sensitivity of nitro energetic compounds using three-dimensional descriptors. Journal of Molecular Graphics and Modelling, 36, pp. 10-19.