CHEMICAL ENGINEERING TRANSACTIONS
VOL. 48, 2016
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Eddy de Rademaeker, Peter Schmelzer

ISBN 978-88-95608-39-6; **ISSN** 2283-9216
DOI: 10.3303/CET1648087

Thermal Stability Predictions for Inherently Safe Process Design Using Molecular-Based Modelling Approaches

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

[a]EPFL, ISIC GSCP Group of Physical and Chemical Safety, Lausanne, Switzerland
[b]Swissi Process Safety, Basel, Switzerland.
thierry.meyer@epfl.ch

For efficient inherently safe design, awareness of all available options at the appropriate decision-making moment is key. Simulations offer both timely availability and large screening possibilities. Therefore computer-aided product design gained in popularity during the recent years and models of various physico-chemical properties were developed. Yet, predictions of safety related data are still limited. Hence, thermal stability derived from Differential Scanning Calorimetry (DSC) is analysed and simulated with two molecular-based modelling approaches: Group Contribution Method (GCM) and Quantitative Structure-Property Relationships (QSPR). Predictive models are developed and evaluated over fitting and predictive abilities for five properties extracted from the DSC curves.

1. Introduction

Any step of an industrial chemical process can potentially involve thermal risks, as most reactions carried out are exothermic, chemicals are often thermally unstable, and operating conditions set to favour high conversion and throughput. In spite of efficient risk assessment methods and implementation of risk mitigation measures, accidents are still occurring. Rather than reducing the risks, a more efficient strategy would be to reduce the hazards: an inherently safer process would present fewer hazards and fewer weak points. To come closer to inherent safety, four basic principles are to be applied (Kletz 2003; CCPS 2009):

- minimize the quantities of hazardous materials handled;
- substitute a hazardous compound for a less hazardous one;
- moderate the potential effects of residual risks;
- simplify the system and avoid additional equipment or features.

These principles offer great potential for safety enhancement, and should be iteratively implemented at various stages of process development. Their impact is optimal when influencing initial choices regarding the selection and development of the process reactions and equipment (CCPS 2009). Moreover, at early design stages, the flexibility is still high and the costs of change low, whereas modifications brought later would incur higher financial costs.

On the other hand, at the preliminary development stages, only limited knowledge and information are available on the final characteristics the process would have and of the chemicals that would be involved. In this context, it is highly valuable to be able to assess beforehand the chemical and physical properties that offer insight in order to make the appropriate developments, and to allow for the conceptual design to be partially performed in silico.

Bringing thermal risk assessment earlier in the design timeline would grant for inherently safer alternatives to be considered at the most appropriate timing. A rigorous thermal risk assessment is part of any chemical process design. Among others characterizations, it requires answering crucial questions regarding the chemical thermal decomposition, if any, at which temperature would it occur, how much heat would be released and at which rate. This information is gathered either from own data, literature, expertise, or experiments. In particular, Differential Scanning Calorimetry (DSC) allows to identify and quantify thermal decompositions characteristics such as potential energy release, triggering temperature, and could serve for thermokinetic data (Sanchirico, 2014). Therefore, DSC experiments will serve here as the basis for these

Please cite this article as: Baati N., Nanchen A., Stoessel F., Meyer T., 2016, Thermal stability predictions for inherently safe process design using molecular-based approaches, Chemical Engineering Transactions, 48, 517-522 DOI:10.3303/CET1648087

thermal stability predictions. Fundamental characteristics of the thermal trail will be examined apart and for each of them predictive models will be produced. The curves will then be simulated from the assembled estimated characteristics in order to obtain an overview of the thermal behaviour.
For predictive models to be applicable at the early development stages, when little information is available and definitive, they should rely on minimal information. So, independently from other considerations concerning the operations, the thermal stability of the compounds will be estimated only from their molecular structures. For this purpose two main methods are widely used: Group Contribution Methods (GCM) and Quantitative Structure-Property Relationship (QSPR). The main difference originates in the way the molecules are defined within these frameworks. For QSPR methods, the molecule is an entity characterized by its constitution and theoretical descriptors derived from its geometry, topology, electronic structure and quantic properties. On the other hand, GCM consider the molecule as the assembly of its constitutional functional groups. Each one contributes to all properties of the molecule, and its contribution is constant if it is comprised in a different structure and depending on its frequency. There are a quantity of group contributions frameworks which define the groups in different manners; in this case, the Marrero-Gani GC+ framework is applied: it consists in Group Contribution enriched with connectivity indices (CI) (Hukkerikar et al. 2012).
In the process safety domain, there have been, in the recent years, efforts to develop predictive models for reactivity hazards and thermal decomposition of various chemicals based on either of these approaches. In particular, DSC derived parameters, and mostly the onset temperature and the decomposition enthalpy have been the focus of several QSPR models (Saraf et al. 2003; Klos et al. 2008; Lu et al. 2011). They are fewer GCM-based models developed for the precisely the same properties (Keshavarz et al. 2009; Lazzús 2012), nonetheless, there were several applications to different thermal properties such as temperature of significant mass losses (Mallakpour et al. 2014) or heat capacities (Pilar et al. 2015). Interestingly, when considering DSC derived parameters, all studies mentioned above focus on particular sets of chemicals with common substructures: i.e. nitroaromatic compounds (Saraf et al. 2003; Keshavarz et al. 2009), aromatic derivatives of carbamic acid (Klos et al. 2008), organic peroxides (Lu et al. 2011) or ionic liquids (Lazzús 2012). The development of models with broader application ranges is rather challenging: a trade-off between the accuracy and the application domain is probably the underlying reason.

2. Modelling process

2.1 Data pre-treatment

A collection of more than 400 DSC experiments of miscellaneous compounds constitutes the initial database for this modeling study. The experiments were performed on a Mettler Toledo DSC1 apparatus, in gold-plated crucibles, closed under inert atmospheres and scanned between 0° and 400°C with a temperature increase of 4°C/min.
The obtained thermograms were subjected to identical treatment using AKTS software (AKTS 2000), namely baseline corrections and fitting by Gaussian like equations to abstract the thermal trail to four characteristics: peak amplitude (A [W/mol]), position of the maximum (Tmax [°C]), peak width (σ [°C] and a factor of asymmetry (AS [-]) to describe if the peak shows either tailing or fronting shapes. Finally, the partial area that represents the reaction enthalpy (ΔHr [J/mol]) is also considered independently.

2.2 Molecular Structure Descriptions

For the generations of the QSPR descriptors, 2D molecular representations were treated with Codessa Pro software (Codessa Pro 2002) to optimize 3D geometries and perform the numerical descriptors calculations. Over a thousand descriptors of their constitution, geometry, topology, and electronic, quantum chemical and thermodynamic properties are generated. A first selection eliminates the descriptors with the lowest variances or with missing values resulting in a working set of approximately 350 descriptors per structure.
Similarly, the 2D molecular structures are also treated with ICAS software (ICAS 2014) to generate Marrero-Gani groups and connectivity indices. To represent all the dataset molecules, 217 groups were necessary.

2.3 Models

Independently of the molecular based method used, the model construction phase can also be adapted from several statistical techniques with varying degrees of complexity. Multi-linear regressions (MLR) are the simplest and most used method, which consists in determining the coefficients for a selection of parameters to fit the observed property. The models are built as MLR following the stepwise regression method. This feature selective method screens among the structural descriptors/groups to find the combination with the highest correlation to the modelled properties by performing a statistical F-test and evaluating the corresponding p-value. The parameters that do not enhance the models are excluded.

It is also important to note that within this study, the dataset is sub-divided before the modeling phase. K-nearest neighbor (k-NN) algorithm is used for classifying the data in four clusters defined by their structural data space. As molecular structures are defined in two different manners, the clustering results are different, and therefore the entire model construction process is performed twice.

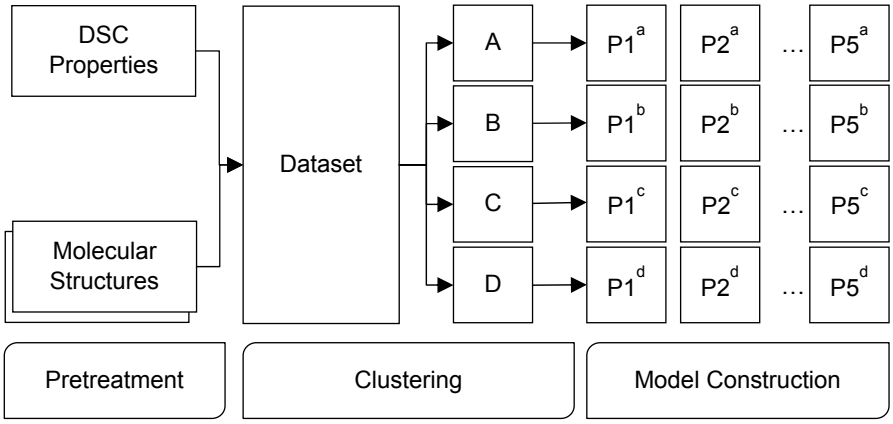

Figure 1: Summarized modelling process

To summarize, figure 1 shows a schematic representation of the modeling process: 5 DSC properties are extracted from all thermograms and molecular structures are described by QSPR and GCM methods (hence the doubled step); by clustering, the dataset is divided in four subsets (A to D); finally, the models are developed for within clusters.

3. Applications examples

Considering that two modeling methods are applied, four clusters created and five properties studied, overall, there are 40 models developed. Among them, 10 are exposed here and summarized in Table 1 to illustrate the obtained results. These models were built for clusters of similar size and distribution, and are thus comparable and allow a parallel discussion. They are characterized by the number of structural descriptors or groups they rely on (*#parameters* in Table 1), the correlation coefficients of the obtained predictions to the observations respectively in the training R^2 and in the validation set R_{cv}^2, and the average absolute relative deviations ARD expressed in percentage of the observed value.

Table 1: Comparative summary of obtained GCM and QSPR models

Method	*Property*	*#parameters*	R^2	R_{cv}^2	ARD_{Tr} %	ARD_{Val} %
GCM	*ΔHr*	18	1	0.91	0	58
	A	18	1	0.81	0	170
	T_{max}	18	1	0.92	0	25
	σ	18	1	0.87	0	28
	AS	18	0.99	0.99	19	51
QSPR	*ΔHr*	10	0.93	0.87	83	68
	A	7	0.83	0.76	49	76
	T_{max}	14	0.97	0.69	5	56
	σ	11	0.93	0.84	35	40
	AS	9	0.90	0.83	54	64

The subsets serving for the construction of the above models comprised common observations including lauroyl peroxide, sucrose, and chloro-2-ethynylbenzene. After estimating the 5 DSC properties with both GCM and QSPR methods, the thermograms of these three compounds were simulated and are shown in Figure 2. This figure shows a comparison of the thermograms obtained from simulations with the corresponding experimental measurements. During the experimental phase, some measurements were replicated, deviations were noted and quantified, and therefore on Figure 2, “tolerance zones” also delimited which represents the most probable DSC curve measured with a ± 5% tolerated error relative to the measured data.

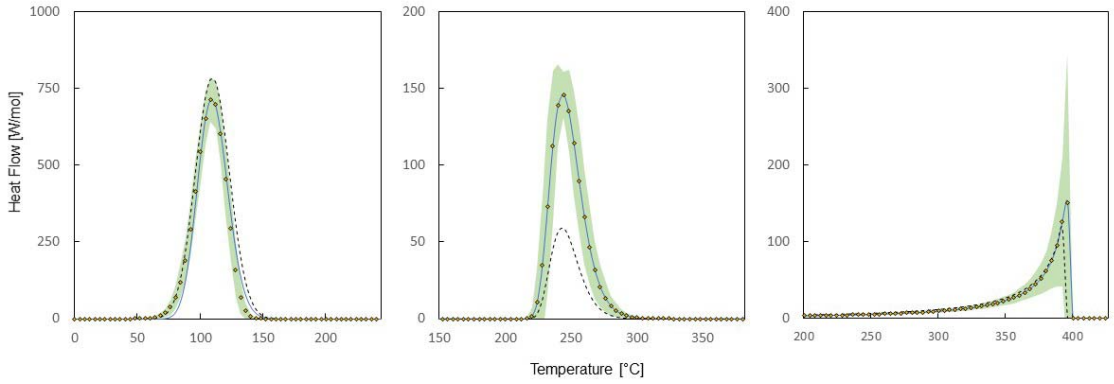

Figure 1: Simulated thermograms with GCM method (full line) and QSPR method (dashed line) in comparison with experimental measurement (diamonds) and "experimental error zone" (shadowed area) – from left to right: lauroyl peroxide, sucrose, and 1-Chloro-2-ethynylbenzene

In order to better illustrate the results, the calculation details for the estimations of the enthalpy of decomposition of Lauryol peroxide ($[CH_3(CH_2)_{10}CO]_2O_2$) with both methods are presented in Table 2. As mentioned above, the models are linear regressions built as

$$P_j^k = \sum_i \alpha_i \cdot x_{ij} + \beta_o \quad (1)$$

i.e. for any compound j, belonging to cluster k, its predicted property P is the sum of all its structural descriptors or group frequency x_{ij} multiplied by their corresponding coefficients α_i, plus an eventual intercept βo. *Table 2* summaries all relevant groups and descriptors necessary to simulate the enthalpy of decomposition of lauryol peroxide. For comparison, the measured enthalpy is ΔHr_{LP} =-322.9 kJ/mol, and the average relative deviations of both simulations are also shown in the table.

Table 2: Detailed calculations for decomposition enthalpy of lauryol peroxide

GCM				
	i	**α_i**	**x_i**	**G_i**	**Definition**
	1	-11.8	18	CH_2	Count of CH_2 in molecule
	2	128	2	CH_2COO	Count of CH_2COO in molecule
	3	32.0	46	H	Count of H in molecule
	4	48.6	4	O	Count of O in molecule
	5	23.8	24	C	Count of C in molecule
	6	-112	18.8	^0X	Kier and Hall Connectivity Index
	βo	0			Y -Intercept
Result					ΔHr_{LP} =-323 kJ/mol
ARD %					0%
QSPR					
	i	**α_i**	**x_i**	**D_i**	**Definition**
	1	-441	0.97	$N_{SB,r}$	Relative number of single bonds
	2	-53.9	11	MNO_{FCmax}	Max atomic force constant
	3	6.65	0.66	RNCS	Relative Negative Charged Surface Area (MOPAC PC)
	4	20.3	0.14	RPCS	Relative Positive Charged Surface Area (Zefirov PC)
	5	342	-2.2	MNO_{ABMO}	Max antibonding contribution of one Molecular Orbital
	6	1209	0.075	FPSA3	Fractional atomic charge weighted Partial Positive Surface Area (MOPAC PC)
	βo	1412			Y -Intercept
Result					ΔHr_{LP} =-283 kJ/mol
ARD %					12.4 %

4. Discussion

From the 23 observations of the subset studied to develop GCM models, 18 serve to adjust the models while 5 are taken out for the validation. All models developed following GCM achieve correlations with $R^2 \geq 0.99$ and besides the asymmetry model, all of them present good accuracy with practically null average relative deviations. The predictive power is not as high as the fitting capability, as the deviations are higher for the validation than the training set, and are probably symptomatic of overfitting. Indeed, when the models are excessively adjusted to the training set, they lose capability of predicting out-of-the-set data. In this case, while the overall tendency is retrieved with good R_{cv}^2 the accuracy is diminished (high deviations). Though the amplitude model shows very high average deviations for the validation set, in reality it is only due to one outlier that when removed leaves the corrected ARD_{val}=66% instead of ARD_{val}=170%. It is noteworthy that the group-contribution models involve certain combinations from an ensemble of 26 groups, mostly 1st order groups and connectivity indices. However, not all of them are relevant for the modeling of molecules as an average of 11 groups describe each molecule. The lauryol peroxide example shows that in that case only 6 groups are necessary to describe its enthalpy for instance. The estimation of ΔHr_{LP} with the GCM model is extremely accurate as the ARD is practically null.

Concerning the QSPR models, a slightly larger dataset is used comprising 33 observations. Overall, these models are weaker than the previous ones: lower correlations coefficients for both training and validation sets, and consequently higher deviations. Both fitting and predictions are limited compared to GCM. The main reason could be a higher sensitivity of QSPR models to over-parametrization. When varying the p-value, the criteria for the inclusion/exclusion of parameters to the models, even minimal variations lead to models with twice as many parameters, presenting high correlations coefficients and null deviations for the training set, but unable to depict the tendency of the validation set, let alone be accurate. For instance, among the potential QSPR models for σ, no models with 13 to 25 parameters were found. There were either regressions combining up to 12 parameters with similar efficiency to the model in Table 1, or over-fitting models with 26 parameters or more presenting very high fittings and poor predictive power. Therefore, despite the high deviations shown in Table 1, these models were selected as the best compromise between efficiency and over-parametrization.

Figure 1 exhibits the graphical comparison of the reconstructed thermograms. The overall appearance shows adequate fittings. Accordingly to the discussion above, GCM present ideal simulations for the three observations shown on Figure 1 and overcome the results of QSPR. Except for the sucrose amplitude that is underestimated, QSPR simulations are correct though, and the DSC reconstructions fall in the “tolerance zone”.

Finally, Table 2 details the calculations of the enthalpy for a particular compound. It is remarkable that most parameters included in the GCM model are rather simple and can be directly extracted from visual inspection of the molecular structure. The only exception is $^0\chi$, the molecular valence connectivity index (Hall & Kier 1986) that is not trivial and needs to be calculated. Interestingly, $^0\chi$, known to be extremely correlated to the molecular surface area - to the extent it is assimilated as an accurate measurement (Karcher & Devillers 1990)- participates in the GCM model while the QSPR model is built mostly with electrostatic and charge-related descriptors which are surface area dependent. On the other hand, the QSPR descriptors selected for modelling ΔHr, and similarly for all modelled properties are more complex and require computational work. That said, once a QSPR descriptors generating software is available for developing such models, the generation of the same descriptors for an extra molecule is not a disputed point. It is more an exportability limitation as the descriptors generation becomes an unavoidable step requiring detailed information on the theoretical assumptions and models for the descriptors calculations. Hence, the overall procedure applicability depends on the availability of the descriptors calculation method and theoretical assumptions.

5. Conclusions

We put forward a method relying on molecular-based approaches to predict thermal stability of a set of thermally reactive miscellaneous chemicals. The efficiency of the models are disparate, nonetheless we overcome to major limitations: in both cases DSC thermograms were recreated from estimated key properties instead of a classical standard two-value characterization of DSC information; moreover, the initial dataset includes structurally diverse chemicals rather than compounds with pre-selected structure similarity. The main objective here is not to state whether GCM or QSPR are more appropriate for DSC simulations, yet for the considered set of observations, GCM models prevailed on all comparative criteria.

These results highlight how molecular based modeling can be applied to properties crucial for thermal risk assessment and this could benefit to inherently safer design. Indeed the predictions offer several advantages that are required to apply inherent safety in the best conditions: an alternative in the absence of practically feasible experiments; a possibility of large screening within limited time and resources, which allows to

determine the best candidate for eventual substitution; if no substitutions are defined, at least the hazards related with the considered chemical are foreseen; and most of all they rely on minimal information and are thus readily available.

Acknowledgments

This work is funded by the Swiss Commission for Technology and Innovation [14711.1 PFIW-IW]. The authors are grateful for the assistance from Novartis SA, AKTS SA (Sierre, Switzerland), and Computer Aided Process Engineering Centre (CAPEC, DTU Lyngby, Denmark).

References

AKTS, AKTS-Thermokinetics (V 3.8), 2000, Advanced Kinetics and Technology Solutions. www.akts.com

Center for Chemical Process Safety (CCPS), 2009. Inherently Safer Chemical Processes - A Life Cycle Approach. 2nd Edition, Hoboken, New Jersey, John Wiley and sons.

CODESSA, Codessa Software (V 1.0 RC2), CODESSA Pro, 2002, University of Florida

Gani and Partners, ICAS (V 17.0) 2014, Computer Aided Process Engineering Centre (CAPEC) www.capec.kt.dtu.dk/Software/

Hall, L.H. & Kier, L.B., 1986. Molecular connectivity and total response surface optimization. Journal of Molecular Structure: Theochem, Elsevier, 134(3-4), pp.309–316.

Hukkerikar, A.S., Sarup, B. Ten Kate, A., Abildskov, J., Sin G. & Gani, R., 2012. Group-contribution + (GC +) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilibria.

Karcher, W. & Devillers, J., 1990. Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Dordrecht, the Netherlands: Kluwe Academic Publishers.

Keshavarz, M.H., Pouretedal, H.R. & Semnani, A., 2009. Relationship between thermal stability and molecular structure of polynitro arenes. Sciences-New York, 16, pp.61–64.

Kletz, T.A., 2003. Inherently Safer Design—Its Scope and Future. Process Safety and Environmental Protection, 81(6), pp.401–405.

Klos, J., Nowicki, P. & Cizmarik, J., 2008. Thermal parameters of phenylcarbamic acid derivatives using calculated molecular descriptors with MLR and ANN: Quantitative structure-property relationship studies. Journal of Thermal Analysis and Calorimetry, 91(1), pp.203–212.

Lazzús, J. A., 2012. A group contribution method to predict the thermal decomposition temperature of ionic liquids. Journal of Molecular Liquids, 168, pp.87–93.

Lu, Y., Ng, D. & Mannan, M.S., 2011. Prediction of the Reactivity Hazards for Organic Peroxides Using the QSPR Approach. Industrial & Engineering Chemistry Research, 50(3), pp.1515–1522.

Mallakpour, S. Hatami, M., Khooshechin, S. & Golmohammadi, H., 2014. Evaluations of thermal decomposition properties for optically active polymers based on support vector machine. Journal of Thermal Analysis and Calorimetry, 116(2), pp.989–1000.

Pilar, R., Pachman, J., Matyáš, R., Honcová, P. & Honc, D., 2015. Comparison of heat capacity of solid explosives by DSC and group contribution methods. Journal of Thermal Analysis and Calorimetry, 121(2), pp.683–689.

Sanchirico R., 2014, A new approach for the reliable estimation of kinetic parameters by means of dynamic dsc
experiments, Chemical Engineering Transactions, 36, 145-150, DOI: 10.3303/CET1436025

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.