Enthalpy of Formation Modeling Using Third Order Group Contribution Technics and Calculation by DFT Method.

Argoub Kadda[1*], Benkouider Ali Mustapha[1], Ahmed Yahiaoui[1], Toubal Khaled[3]; Djebar Hadji[2]

[1] Laboratory of Organic Chemistry, Macromolecular and Materials, Department of Chemistry, University Mustapha Stambouli of Mascara, BP 763, 29000 Mascara, Algeria.
[2] Modeling and Calculation Methods Laboratory, Department of Chemistry, Dr. Moulay Tahar University of Saida, 20000, Saida, Algeria
[3] Laboratory of Applied Organic Synthesis, Department of Chemistry, Faculty of Sciences, University of Oran 1 Ahmed Ben Bella, 31000, Oran, Algeria
argoubkadda@yahoo.fr

Abstract

Recently, with the development of calculators and numerical tools, quantum computations to explore the electronic, structural and dynamic properties of matter without resorting to experimental knowledge have seen increasing development. Thus, it is possible to perform ab-initio calculations with increasing precision and for increasingly larger systems. In the scientific literature, papers using ab-initio quantum computation for the prediction of formation enthalpies is more and more numerous. The aim of this paper is to develop a theoretical method to calculate standard enthalpy of formation in gas stat for organic compounds using group contribution technics (third-order group contribution method). For the establishment of this method, 750 molecules are used. In parallel with group contribution methods, this paper presents another approach to calculate gas-state formation enthalpies based on DFT method. The calculation involved 30 molecules with at least one ring from C3 to C13. Finally, DFT and group contribution results are compared.

Keywords: Group contribution method; DFT; Organic compound; B3lyp, 6-31g.

1. Introduction

In today's chemical processes hundreds of thousands of components are used. The Chemical Abstracts Service registry lists 56 million substances [1], but many of these are only of scientific interest. Process designers need to know some basic chemical properties of the components and their mixtures. Experimental measurement is often too expensive. Predictive methods can replace measurements if they provide sufficiently good estimations [2]. The estimated properties cannot be as precise as well made measurements, but for many purposes the quality of estimated properties is sufficient. Predictive methods can also be used to check the results of experimental work.

The standard enthalpy of formation or standard heat of formation of a compound is the change of enthalpy during the formation of 1 mole of the substance from its constituent elements, with all substances in their standard states.

Enthalpy is an important parameter, which can be useful to evaluate energy especially for some thermal machine like gas turbine. A gas turbine is a combustion engine that can convert gas or other liquid fuels to mechanical energy [3-8]. So developing a theoretical method to estimate formation enthalpy especially for gas stat became a necessary for ameliorating the gas turbine performance [9-12].

The Methods for determining the $\Delta_f H^\circ_{gas}$ values of pure components can roughly be divided into two important classes: quantum mechanical approach [13-18] and group contribution approach [19-24]. Quantum mechanical methods provide generally reliable predictions for very small molecules [25-27]. However, these methods require high level theoretical calculations and significant computing power. By using quantum mechanical approach, several authors viz. Keshavarz et al. [28], Mathieu and Simonetti [13], Rice et al. [14] and Politzer et al. [16] proposed to estimate the formation enthalpies. Their methods showed a low deviation from experimental data only for a small data set of organic compounds. They estimated enthalpies of formation in the gas phase by using different density functional theory methods. For computation of sublimation enthalpies, Keshavarz et al. [17] and Rice et al. [14] have used the molecular surface area, total average potentials and the positive and negative variance of the surface electrostatic potential. As for other authors they modified the van der Waals electrostatic surface potentials of individual molecules for computation of the sublimation enthalpy.

For group contribution method is used to predict properties of pure components and mixtures by using group or atom properties. This reduces the number of needed data dramatically. Instead of needing to know the properties of thousands or millions of compounds, only data for a few dozens or hundreds of groups have to be known. In the literature, several methods for estimating $\Delta_f H^\circ_{gas}$ with good accuracy have been proposed to fill the gaps between existing compounds and their experimental values of formation enthalpies [17, 19, 23, 25, 29-31]. For gas and liquid phases, it is now possible to calculate formation enthalpies values for some compounds with reasonable accuracy [32].

The objective of this work is to develop a simple accurate model for the prediction of enthalpies of formation at gas state from the molecular structures of pure organics compounds with a wide range of applicability (C, H, N, O), and to show a detailed analysis of its performance compared to previous models. The development of a thermodynamic model for the estimation of $\Delta_f H^\circ_{gas}$, is based on the reliability of the experimental data. Building a good model requires a large database with a variety of chemical functions.

The Marrero and Gani scheme [32] is used (third order group contribution method). To improve the reliability and accuracy of proposed model, especially in the case of C H N and C H N O compounds, new structural groups are defined and included in the Marrero and Gani [32] scheme based on the analysis of the predicted results and molecular structure compounds. In another hand enthalpies of formation at gas state are calculated using GAUSSIAN 09 [33]. The results from the two methods are compared.

2. Experimental Data

The experimental values of $\Delta_f H^\circ_{gas}$ presented by National Institute of Standards and Technology (NIST) [34] were collected and implemented with those recently published [28, 35, 36] to provide the reliable data set. The data set used in the present study is comprised of 750 experimental data points including 238 CH, 155 CHO, 138 CHN and 219 CHNO compounds. All values stored in the data set, range in $\Delta_f H^\circ_{gas}$ values from -1285 up to 705 kJ/mol. The $\Delta_f H^\circ_{gas}$ values are expressed in kJ/mol which facilitated the comparison of proposed model results with those of other research.

The quality of a predictive method depends on a variety of factors such as quality of data set used in its development and validation. For this reason, an analysis of experimental data was performed to check inaccuracies/uncertainties of the data, by consulting mainly National Institute of Standards and Technology (NIST) [34], other works and compilations. When one experimental measurement is available in the NIST database, this value is considered automatically. CAS Registry Numbers and $\Delta_f H^\circ_{gas}$ values for each compound are presented as supplementary material. (See appendix D)

2.1 Model Development

Having defined the data set of $\Delta_f H^\circ_{gas}$, the collection of structural groups (molecular fragments) is considered. The fragmentation of the molecular structures was performed by an automatic program using Marrero and Gani method [32]. For more details on the cutting technique we recommend these works [32, 37, 38] . The chemical structures of all

considered compounds were used. It was found that the Marrero–Gani's first-order groups do not describe all compounds studied. For this reason, new first-order groups have been introduced and presented as supplementary material (See appendix C).

The random choice of a validation set may distort the performance of the model if a contribution is regressed on a single data point or two (or even three) and may exclude some model parameters (group-contributions) from the training step which affects the application range [39, 40]. So, if group contribution value is generated from less than three compounds, this group contribution would certainly be more inaccurate than others.

For this purpose, in this work the data set was divided into two sub data sets using semi-randomly choice [37, 38]. Indeed, if a compound is selected and if this compound is described by one of the irrelevant groups; group having their contribution value generated by less than three compounds, this compound is removed from the test set. This is important to avoid the lack of some necessary data for training step and to allow better regression of some model parameters. About 80% of the main data set are selected for training set (600 compounds) and all other (20%) are used for the test set (150 compounds).

For this purpose we have developed a MATLAB program. It can divide in an automated way the data base into training and test set. This program uses the SMILES code. It is a semi random division following these steps:
(1) Verification of the groups (descriptors) which describes the molecules.
(2) Calculates occurrences of each group for all the databases. Then we can know how many times the groups (descriptors) are mentioned.
(3) If a group is mentioned less than three times, molecules correspond to this group, are included in the learning base.
(4) If the group is mentioned more than three times, molecules correspond to this group, are included in the learning or test base in a random choice.

The combination of the third and the fourth criteria gives the semi-random division. In this work, a simple practical multivariable linear model is proposed. Several mathematical functions were tested to represent the $\Delta_f H^\circ_{gas}$, The best mathematical form was found to be a simple linear form. The model obtained is as follows:

$$\Delta_f H^\circ_{gas} \left(\frac{kJ}{mol} \right) = \sum_{i=1}^{N} n_i A_i \qquad (1)$$

where A_i is the contribution of first-, second-, or third-order group of type i that occurs n_i times. To estimate the model parameters (contributions A_i) a simultaneous regression is considered by used all contributions of first-, second- and third-order groups in the training phase of the proposed model. This avoids that the higher-order terms absorb the errors of the lower-order terms. These parameters are obtained by minimization the sum of the squares of the estimation errors using Levenberg–Marquardt algorithm. Having estimated the values of $\Delta_f H^\circ_{gas}$ for all the compounds in the dataset using Eq. (1), it was found that the average absolute deviations (AAD) for all compounds are relatively large; there is a fraction of compounds, which have large estimation errors. By examining these last compounds, it was found that these compounds belong to the following chemical classes: cyclic alkanes, acyclic and cyclicpolynitro, six-member and five-member ring, quinolines and compounds with some specific molecular

fragments. After analyzing the molecular structures of these compounds having large estimation errors, new structural groups are defined and included in the Marrero–Gani's groups as additional new third-order groups in order to increase the accuracy and reliability of the proposed model. These new groups provide more structural information and allow distinguishing between similar compounds having one or more functional first-order groups in common [37, 38]. A list of new groups, along with sample assignments and group occurrences, is presented in appendix C (See supplementary material). Finally, 314 groups were found to predict the $\Delta_f H^\circ_{gas}$ values, viz. 115 first-order, 77 second-order, 36 third-order groups and 83 new groups. The first-, second-, third-order and new groups, their contributions to the $\Delta_f H^\circ_{gas}$ and their number of occurrences in the dataset are presented in appendix C. As well, the number of occurrences of the groups in each compound used in this work is presented. The application of the proposed model is illustrated through an example in Appendix B.

The proposed model does not distinguish $\Delta_f H^\circ_{liq}$ of several isomers in particular stereoisomers, for example, cis and trans isomers. However, the proposed method is able to distinguish between some isomers like branched alkanes with methyl group and branched aromatic compounds.

In this work, we have not added new functional groups to distinguish the different isomers in order to respect the structural groups of the Marrero and Gani method [32] in their standardized form and thus facilitate the integration of our method in software and application.

3. DFT Calculation of Formation Enthalpies in the Gaseous State

In parallel with group contribution methods, this paper presents another approach to predict gas-state formation enthalpies based on an ab initio quantum molecular energy calculation. The calculation involved 30 molecules with at least one ring from C3 to C13. These molecules have been retained because they present some problems with the group contributions techniques.

3.1 Geometric Optimization with Frequency Calculation

Before starting the calculation of the formation enthalpy of the chosen organic compounds, the molecule must be optimized. This operation involves entering the input file and then launching the optimization procedure by choosing the "optimization" option from the choices given in "job type". After having optimized the structure, one must launch the second operation to have the thermodynamic parameters by using the "frequency" option. Note that you must use the same method and the same basic set for optimizing and calculating frequencies.

In our work, we realized the choice "Optimization + Frequency" in the interface of the software. This choice consists in performing in parallel the geometric optimization of the molecule at the same time as the frequency calculation in the same operation. This technique makes it possible to optimize the

The molecular structure and calculate the frequencies. In this work B3LYP method is used with 6-31G basis set. The stability of the molecule can be easily controlled by checking for imaginary (negative) frequencies. The imaginary frequencies indicate instability of the molecular geometry. Negative frequencies indicate instability in the molecule. A stable molecule should not have imaginary frequencies.

3.2 Calculation of the Formation Enthalpy at 298.15 K

After entering the input file in the G09 [33] software and finishing the quantum computation (output file appearance in G09), the first task is to confirm that we have attained the convergence by the frequency check. If one of the calculated frequencies is negative (imaginary frequency) it means that we have not found the energy corresponding to the most stable conformation and that this energy is for a local minimum instead of the global minimum.

The output file provided by G09 gives the formation enthalpy values of the compounds with respect to the ab-initio reference state which consists of atomic nuclei and separated electrons, under zero pressure and at the temperature of 0 K. This value is expressed in Hartree per molecule (1 Hartree per molecule = 2625.50 kJ/mol).

The transition from the ab-initio reference state to the thermochemical reference state consists of subtracting from the enthalpy obtained by ab initio for the molecule studied the enthalpy obtained equivalently for its elementary constituents, then adding the term to change from a monoatomic reference state to a molecular reference state [5]. Then it is necessary to calculate the atomic parameters C_i then the calculation of the enthalpy of formation in the standard state in the thermochemistry, with the standard pressure $P° = 101325$ Pa and the temperature T = 298,15K. This calculation is given by the following equation:

$$\Delta_f H^\circ = H^{Al} - \sum v_i \left(H_i^{Al} - \Delta_f H_i^\circ \right) = H^{Al} - \sum v_i C_i \qquad (2)$$

H_i^{Al} : Ab-initio Enthalpy of the element i
$\Delta_i H_i^\circ$: Standard formation enthalpy in perfect monoatomic gas state
v_i : Stoichiometric coefficient
C_i : Atomic parameter

Atomic parameters can be defined by ab-initio calculations or determined using the standard monatomic gas (gas perfect) formation enthalpies found in the literature. Another technique is to adjust these parameters to experimental data by minimizing the following function.

$$J = \sum \left[\Delta_f H_j^\circ - \left(H_j^{Al} - \sum_i v_{ij} C_i \right) \right]^2 \qquad (3)$$

Table 1 shows the values of the atomic parameters of carbon and hydrogen determined by enthalpies of formation of atoms and those which are adjusted according to the least squares method [16].

Table 1. Result of optimization of the atomic parameters.

A_i	Defined values	Adjusted values
C	37.773922	38.113836
H	0.4979950	0.5825820

As part of this work, tests were carried out to calculate the formation enthalpy in the state of gas under the standard conditions of pressure and temperature, using the two techniques for calculating the atomic parameters; purely ab-initio and adjusted to experimental data. The best results found are those by the use of adjusted parameters.

2.2 Results and Discussion
First Method

The performance of the proposed model for prediction of $\Delta_f H_{gas}^{°}$ values was evaluated, analyzed, compared to previous models and discussed. Generally, in the case of a large data set it is necessary to examine in detail the results of the proposed model. The statistical performance indicators used in this work, viz. coefficient of determination (R^2), standard deviation (SD), mean square error (MSE) and average absolute deviation (AAD) are defined in Appendix A.

After having developed a new method for the prediction of gaseous formation enthalpies, our model must then be validated and tested the predictive power. In a first step, the proposed model is applied for the estimation of formation enthalpies of the molecules used in the development of the model (training set). Then the proposed model is applied to new molecules that have not been used in its implementation to properly test its predictive power and robustness.

The different statistical indicators namely AAD, SD, MSE, and R^2 are displayed as follows; 3.33, 3.06, 4.52 and 0.9997, respectively in the case of training data. The indicators are 4.50, 4.04, 6.04 and 0.9992 respectively in the case of the test set. Table 2 shows the maximum deviations (D max) in the two learning and test sets as well as for all the data in addition to the statistical indicators mentioned above. Maximum deviations remain reasonable for both sets.

Table 2. Performance statistics of the developed model.

Data sets	Number of molecules	kJ/mol			
		AAD	SD	MSE	ARD (%)
Training set	600	3.33	3.06	4.52	0.55
Test set	150	4.50	4.04	6.04	2.00
Total set	750	3.57	3.31	4.86	1.00

In order to better highlight the performances of the proposed model, the prediction errors of the $\Delta_f H_{gas}^{°}$ of two databases are classified according to four error intervals namely 0-5 kJ/mol, 5-10 kJ/mol, 10-15 kJ/mol and > 15 kJ/mol (Table 2). A large proportion (72.40%) of the molecules have prediction errors between 0 and 5 kJ/mol.

From these results it can be seen that by using the developed method there is a fraction of 543 compounds witch have deviations greater than 0 kJ/mol but less than 5 kJ/mol. There are also 164; 40; and 3 molecules with an AAD between 5 and 10; 10 and 15, and less <20 kJ/mol respectively. Having examined the molecular structures of these compounds, it was found that they belong to the four classes and there is no relation between their chemical structures. Several reasons may explain these large deviations. Possible explanations include; a review of experimental values of $\Delta_f H_{gas}^{°}$ from NIST data base for some compounds reveals that there is a discrepancy between certain values of $\Delta_f H_{gas}^{°}$, and there are insufficient experimental values for some compounds, which have been published only once. Although the number of input parameters of the proposed method is significantly lower than those of other methods, this new method is more accurate than previous methods.
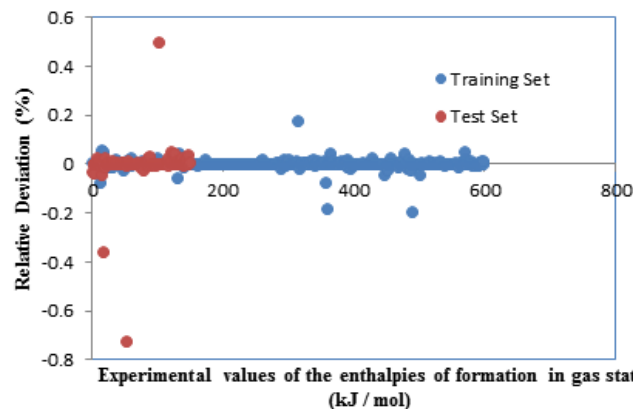

Figure 1. Relatives deviations between the calculated and experimental $\Delta_f H_{gas}^{°}$ values as a function of experimental $\Delta_f H_{gas}^{°}$ values.

The relative deviations values of the developed model for training and test set as a function of experimental $\Delta_f H_{gas}^{°}$ values are shown in Figure 1. The relatives deviations are between -0.20 (Nitroxoline) and 0.17 (Butanenitrile, 2-methyl-) for the training set. For the test set relatives deviations are between -0.73 (1-Butene) and 0.50 (Homocubane-4-carboxylic acid). For the global data set relatives deviations are between -0.73 and 0.50. The mean absolutes deviation for training; test and the total data set are presented in table 2.
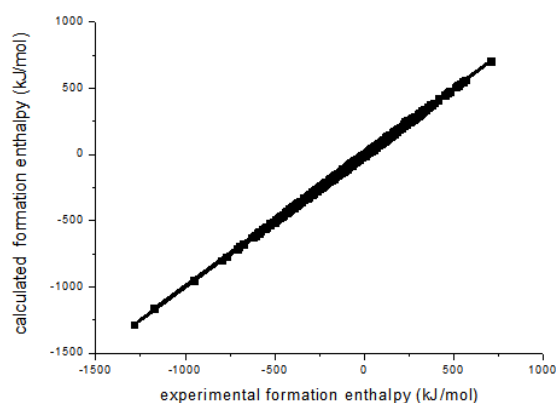

Figure 2. Relationship between the experimental and estimated (by GCM) values of the formation enthalpy of 750 molecules in the gas state.

Table 3. Performance statistics of the developed model.

Methods	Phase	Data set	kJ/mol		
			AAD	SD	D$_{max}$
NGCM	gas	750	3,57	3,31	18,53
M. Benson	gas	799	4.10	7.76	56.60
M. Joback	gas	378	8.40	18.00	*
M. Constantinou	gas	373	3.71	*	*
M. Albahri	3 states	584	11.38	*	1276.6
M. Bagheri	3 states	1694	56.98	*	580.3

NGCM: the new group contribution method for calculation of $\Delta_f H_{gas}^{°}$.

Comparisons have been made between the proposed model and the best existing GC models for estimation the $\Delta_f H_{gas}^{°}$ in gas state. Also, competing models developed to predict $\Delta_f H_{gas}^{°}$ of compounds in in the three states of matter, are considered. Among these models, we can mention those of Benson et al. [41], Joback et al. [42],

Constantinou et al.[43], Albahri et al. [44] and Bagheri et al. [45].

Table **3** shows these comparisons, where statistical indicators as well as AD_{max} are given for each model. Taking into account the size and type of the data, the results achieved by our model are significantly better than those obtained by other models. The database used in this work is the largest compared to those used by other previous models, with the exception of the Bagheri et al database [45]. This increases the application range of the proposed model. It should be noted that the database proposed by Bagheri et al.[45] was taken from the DIPPR 801 database [46], where data are not all experimental (stored as experimental, predicted, unknown, or smoothed) and unfortunately the authors did not mention any information on data quality.

In comparison to the competing models with a large database, the proposed model is more accurate because it displays both a very low AAD and AD_{max}. Benson et al. [41] have a good AAD but these models are applicable only on some classes of compounds, which is a disadvantage. In addition, these methods show a significant AD_{max} values compared to the proposed model.

Second Method

The calculation involved 30 molecules is gaseous stat with at least one ring from C3 to C13. After geometry optimizations for the 30 molecules with Gaussian 09, and obtaining calculated formation enthalpies, the reference is changed to have the standard formation enthalpy at 101325 Pa and 298.15 K.

To analyze these results, the various statistical indicators (absolute average errors, standard deviations, etc.) are calculated. The results are shown in Table 4. For comparison, the statistical parameters of the 30 molecules are recalculated applying the group contribution model. (See Table 4.)

Table 4. Standard formation enthalpy prediction results for gas phase according to DFT and GCM.

	kJ/mol				
Data set	AAD	SD	MSE	R^2	D_{max}
DFT	14.22	12.34	19.65	0.98787	42,38
NGCM	19.46	26.19	32.27	0.97519	87.04

DFT: density functional theory; NGCM: the new group contribution method for calculation of $\Delta_f H^{\circ}_{gas}$

These statistical indicators are recalculated for the 30 molecules (using the NGCM to calculate $\Delta_f H^{\circ}_{gas}$) to compare with those of DFT method. The statistical indicators AAD, SD, MSE and R^2 for the DFT method are 14.22, 12.34, 19.65 kJ/mol and 0.98787 respectively. And For the NGCM, statistical indicators are 19.46, 26.19, 32.27, 0.97519 respectively.

Figure 3 shows an affinity between the values of experimental formations enthalpies and those calculated by the DFT. In this figure, we find quasi-perfect regression between the experimental and calculated $\Delta_f H^{\circ}_{gas}$, which amounts to saying that the computation is strongly related to the experimental values. Indeed, we obtained a linear correlation coefficient $R^2 = 0.98787$.
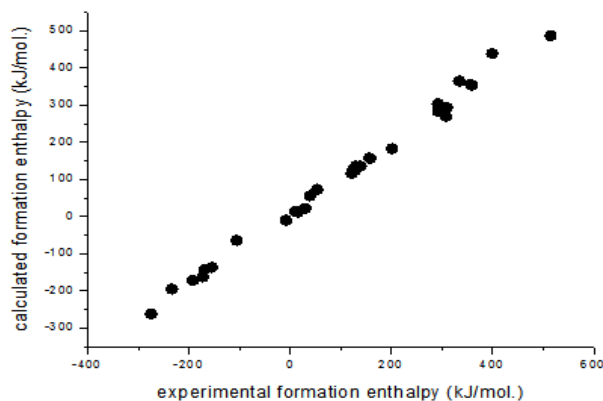

Figure 3. Relationship between the experimental and estimated (by DFT) values of the formation enthalpy of 30 molecules in the gas state

For more results, Appendix E in Supplementary Material gives for each molecule the experimental values (kJ/mol), the values calculated by Gaussian 09 in Hartree per molecule and in kJ/mol, the calculated $\Delta_f H^{\circ}_{gas}$ values using the new group contribution method.

4. Conclusions

The experimental values of $\Delta_f H^{\circ}_{gas}$ can be obtained from traditional combustion calorimetry, but also from enthalpies of reactions measured by reaction calorimetric, as well as from equilibrium studies as a function of temperature. It is usually determined directly for one phase (liquid, gas, or crystal) and the values for other phases are obtained by adding or subtracting enthalpies of vaporization or sublimation [47]. However, obtaining experimental data of this property is not always available, because the measurements of these data are expensive, time consuming and sometimes risky to handle. Even with the availability of the experimental equipment, it is not easy to get very correct values in certain instances, because chemicals often degrade during the measure process [38].

A new third order groups contributions method has been developed for the prediction of the formation enthalpy of organic compounds, in the gaseous state, containing the elements C, H, N and O. The proposed method is simple to implement, it requires only the knowledge of the developed formula of a given compound to apply it without resorting to other thermodynamic properties. A database of 750 molecules was used during the development of the method. An 80% of the total database (training set) was used for the implementation of the model and the rest of the molecules (test set) are used to test and validate the proposed model. The proposed model displays acceptable statistical indicators with an AAD and SD of 3.57 kJ/mol and 3.31 kJ/mol, respectively. From the comparative study it was found that the developed method is more accurate than the method of Benson et al.

Density Functional Theory is an entirely predictive method, used to define the most stable conformation. The DFT can be used also to calculate the parameters corresponding to these conformations. In this work, we presented the calculation details of the standard formation enthalpy of cyclic compounds, in the gaseous state using B3LYP method and 6-31G basis set. The results found are relatively accurate. The deviations AAD, SD, MSE and R^2 are 14.22, 12.34, 19.65 kJ/mol and 0.98787, respectively.

The comparative study between DFT and the new group contribution method developed during this work shows that the results are very close. DFT is a good method for

estimating formation enthalpy, which is very accurate for simple molecules. But once the size of the molecule increases the deviations tend towards enormous values and the calculation time becomes very important.

The ab-initio calculations make it possible to capitalize the theoretical knowledge around the property to be studied and they present considerable improvement prospects compared to the group contributions methods.

Appendix A. Statistical performance indicators used in this work are presented as follows:

AAD: Average absolute deviation:

$$AAD = \frac{1}{N}\sum_{j=1}^{N} |\left(X_j^{exp} - X_j^{pred}\right)|$$

ARD: Average relative error:

$$ARD \% = \frac{1}{N}\sum_{j=1}^{N} |\left(\frac{X_j^{exp} - X_j^{pred}}{X_j^{exp}}\right)| * 100$$

SD: Standard deviation:

$$SD = \sqrt{\frac{1}{N}\sum_{j=1}^{N} \left(X_j^{exp} - X_j^{pred}\right)^2}$$

MSE: Mean square error

$$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(x_{exp} - x_{pred}\right)^2}$$

Order of fragmentation	Occ.	Cont.
First-order of fragmentation		
CH_2	01	-20.84
aCH	10	15.22
aC except as above	01	2.90
aC-NH	01	44.58
OCH_2CHOH	01	-358.88
Second-order of fragmentation	**Occ.**	**Cont.**
$CH_m(OH)CH_n(NH_p)$ (m,n,p in 0..2)	01	-4.48
Third-order of fragmentation	**Occ.**	**Cont.**

Occ: Occurrences; Cont: Contribution values.

References

[1] The Chemical Abstracts Service Chemical Registry System [database on the Internet] 2019. Available from: https://www.acs.org/content/acs/en.html.

[2] Babalola FU, Susu AA. Model development of a suitable equation of state for multicomponent multiphase systems: application to crude oil phase stability requirements. *Int. J. Thermodyn*. 2018;21:111-118.

[3] Brus G, Komatsu Y, Kimijima S, Szmyd J. An analysis of biogas reforming process on Ni/YSZ and Ni/SDC catalysts. *Int. J. Thermodyn*. 2012;15:43-51.

Linear regression coefficient (R^2)

$$R^2 = 1 - \left(\frac{\frac{1}{N}\sum_{j=1}^{N}\left(X_j^{exp} - X_j^{pred}\right)^2}{\sum_{j=1}^{N}\left(X_j^{exp} - \mu\right)^2}\right)$$

with:
x^{exp} : Experimental value
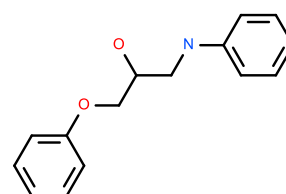X^{pred} : Predictive value.
μ : Average
N : Number of molecules

Appendix B. Illustrative examples:

In the following illustrative example, we present the application of the proposed method for the prediction of enthalpy of formation in gas phase.

Example: Prediction of the formation enthalpy of: 2-Propanol, 1-phenoxy-3 (phenylamino)- (CAS N°: 16112-55-3).

The experimental value of the enthalpy of formation is: -185.00 kJ/mol. The decomposition into groups is given in the following table:

[4] Declaye S, Dumas X, Ferrand L, Lemort V. Waste heat recovery by means of Organic Rankine Cycle (ORC) system coupled with two-phase closed thermosyphons. *Int. J. Thermodyn*. 2017;20:81-89.

[5] Manfrida G, Secchi R. Performance prediction of a small-size adiabatic compressed air energy storage system. *Int. J. Thermodyn*. 2015;18:111-9.

[6] Yasmin M, Gupta M. Thermodynamical Study of Alcoholic Solutions of Poly (ethylene glycol) Diacrylate and Poly (ethylene glycol) Dimethacrylate. *Int. J. Thermodyn*. 2012;15:111-117.

[7] Mousavi SM. Numerical study of entropy generation in the flameless oxidation using large eddy simulation model and OpenFOAM software. *Int. J. Thermodyn*. 2014;17:202-208.

[8] Abolpour B, Shamsoddini R. A novel scheme for predicting the behaviors of liquid and vapor phases of water using the ideal gas theory. *Int. J. Thermodyn*. 2018;21:174-178.

[9] Atakan B. Gas turbines for polygeneration? A thermodynamic investigation of a fuel rich gas turbine cycle. *Int. J. Thermodyn*. 2011;14:185-192.

[10] Oyedepo S, Kilanko O. Thermodynamic analysis of a gas turbine power plant modelled with an evaporative cooler. *Int. J. Thermodyn*. 2014;17:14-20.

[11] Lazzaretto A, Manente G. A new criterion to optimize ORC design performance using efficiency correlations for axial and radial turbines. *Int. J. Thermodyn*. 2014;17(3):192-200.

[12] Qun Z, Hai Z, Lanxin S. Effects of water droplets on the numerical simulation of a complete gas turbine. *Int. J. Thermodyn*. 2018;21(1):7-14.

[13] Mathieu D, Simonetti P. Evaluation of solid-state formation enthalpies for energetic materials and related compounds. *Thermochimica Acta*. 2002;384(1-2):369-75.

[14] Rice BM, Pai SV, Hare J. Predicting heats of formation of energetic materials using quantum mechanical calculations. *Combust. Flame*. 1999;118(3):445-58.

[15] Kim CK, Lee KA, Hyun KH, et al. Prediction of physicochemical properties of organic molecules using van der Waals surface electrostatic potentials. *J. Comput. Chem*. 2004;25:2073-2079.

[16] Politzer P, Murray JS, Edward Grice M, Desalvo M, Miller E. Calculation of heats of sublimation and solid phase heats of formation. *Molecular Physics*. 1997;91:923-8.

[17] Keshavarz MH, Zamani M, Atabaki F, Monjezi KH. Reliable approach for prediction of heats of formation of polycyclic saturated hydrocarbons using recently developed density functionals. *Comput. Theor. Chem*. 2013;1011:30-36.

[18] Keshavarz MH. Simple procedure for determining heats of detonation. *Thermochimica Acta*. 2005;428(1-2):95-9.

[19] Keshavarz MH. Prediction of the condensed phase heat of formation of energetic compounds. *J. Hazardous Materials*. 2011;190(1-3):330-344.

[20] Albahri TA, Aljasmi AF. SGC method for predicting the standard enthalpy of formation of pure compounds from their molecular structures. *Thermochimica Acta*. 2013;568:46-60.

[21] Salmon A, Dalmazzone D. Prediction of Enthalpy of Formation in the Solid State (at 298.15 K) Using Second-Order Group Contributions—Part 2: Carbon-Hydrogen, Carbon-Hydrogen-Oxygen, and Carbon-Hydrogen-Nitrogen-Oxygen Compounds. *Journal Physical Chemical Reference Data*. 2007;36:19-58.

[22] Salmon A, Dalmazzone D. Prediction of enthalpy of formation in the solid state (at 298.15 k) using second-order group contributions. Part 1. Carbon-hydrogen and carbon-hydrogen-oxygen compounds. *Journal Physical Chemical Reference Data.* 2006;35:1443-1457.

[23] Cohen N. Revised group additivity values for enthalpies of formation (at 298 K) of carbon–hydrogen and carbon–hydrogen–oxygen compounds. *Journal of Physical and Chemical Reference Data*. 1996;25:1411-1481.

[24] Domalski ES, Hearing ED. Estimation of the thermodynamic properties of C-H-N-O-S-halogen compounds at 298.15 K. *Journal of Physical and Chemical Reference Data*. 1993;22(4):805-1159.

[25] Hukkerikar AS, Meier RJ, Sin G, Gani R. A method to estimate the enthalpy of formation of organic compounds with chemical accuracy. *Fluid Phase Equilibria*. 2013;348:23-32.

[26] Wang K, Brewster M. Thermodynamic Behavior of Water from Soft-Cell Theory. *Int. J. Thermodyn*. 2011;14(1):1-9.

[27] Singh M. Thermodynamics of Philicphobic Interaction Shift in Aqueous Tweens 20 to 80. *Int. J. Thermodyn*. 2011;14(3):135-46.

[28] Ghaieni HR, Tavangar S, Ebrahimzadeh Qhomi MM. Simple correlations for calculating NHTPB enthalpy of formation through molecular structures. *Multidiscipline Modeling Materials Structures*. 2019;15(1):258-64.

[29] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria*. 2012;321:25-43.

[30] Gharagheizi F. Prediction of the standard enthalpy of formation of pure compounds using molecular structure. *Aust J. Chem*. 2009;62(4):376-81.

[31] Zhang Y. An improved QSPR study of standard formation enthalpies of acyclic alkanes based on artificial neural networks and genetic algorithm. Chemom. Intell. Lab. Syst. 2009;98(2):162-72.

[32] Marrero J, Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*. 2001;183:183-208.

[33] Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 09. Wallingford, CT2016.

[34] NIST Chemistry WebBook, NIST Standard Ref-erence Database Number 69, National Institute of Standards and Technology,Gaithersburg, MD, 2005 http://webbook.nist.gov (retrieved 25.04.13). [database on the Internet]2019.

[35] Mielczarek DC, Nait Saidi C, Paricaud P, Catoire L. Generalized Prediction of Enthalpies of Formation Using DLPNO-CCSD (T) Ab Initio Calculations for Molecules Containing the Elements H, C, N, O, F, S, Cl, Br. *J. Comput. Chem*. 2019;40(6):768-93.

[36] Pimenova S, Lukyanova V, Ilin DY, Druzhinina A, Dorofeeva O. Standard enthalpies of formation of dicyclopropyldinitromethane and tricyclopropylmethane. *J. Chemical Thermodynamics*. 2019;132:316-21

[37] Guella S, Argoub K, Benkouider AM, Yahiaoui A, Kessas R, Bagui F. Artificial Neural Network-Group Contribution Method for Predicting Standard Enthalpy of Formation in the Solid State: C–H, C–H–O, C–H–N, and C–H–N–O Compounds. *Int. J. Thermophys*. 2015;36:2820-32.

[38] Argoub K, Benkouider AM, Yahiaoui A, Kessas R, Guella S, Bagui F. Prediction of standard enthalpy of formation in the solid state by a third-order group contribution method. *Fluid Phase Equilibria*. 2014;380:121-7.

[39] Phifer JR, Cox CE, da Silva LF, et al. Predicting the equilibrium solubility of solid polycyclic aromatic hydrocarbons and dibenzothiophene using a combination of MOSCED plus molecular simulation or

electronic structure calculations. Molecular Physics. 2017;115(9-12):1286-300.

[40] Ackermann T. SW Benson: Thermochemical Kinetics. Methods for the Estimation of Thermochemical Data and Rate Parameters. John Wiley & Sons, Inc., New York, 1968. XII und 223 Seiten, 4 Abbildungen. Preis: 94 s. *Zeitschrift für Physikalische Chemie*. 1969;73(2):241-.

[41] Ackermann, T., SW Benson: Thermochemical Kinetics. Methods for the Estimation of Thermochemical Data and Rate Parameters. John Wiley & Sons, Inc., New York, 1968. XII und 223 Seiten, 4 Abbildungen. Preis: 94 s. *Berichte der Bunsengesellschaft für Physikalische Chemie*, 1969. 73: p. 241-241.

[42] Joback, K.G. and R.C. Reid, Estimation of pure-component properties from group-contributions. *Chemical Engineering Communications*, 1987. 57(1-6): p. 233-243.

[43] Constantinou, L. and R. Gani, New group contribution method for estimating properties of pure compounds. *AIChE Journal*, 1994. 40(10): p. 1697-1710.

[44] Albahri, T.A. and A.F. Aljasmi, SGC method for predicting the standard enthalpy of formation of pure compounds from their molecular structures. *Thermochimica Acta*, 2013. 568: p. 46-60.

[45] Bagheri, M., et al., Molecular modeling of the standard state heat of formation. *Energy Conversion Management*, 2013. 65: p. 587-596.

[46] DIPPR, ed. by Design Institute for Physical Properties (DIPPR) (American Institute of Chemical Engineers (AIChE), New York, 2014.

[47] Verevkin, S., et al., New group-contribution approach to thermochemical properties of organic compounds: hydrocarbons and oxygen-containing compounds. *Journal Physical Chemical Reference Data*, 2013. 42(3): p. 033102.