A QSPR STUDY OF NORMAL BOILING POINT OF ORGANIC COMPOUNDS (ALIPHATIC ALKANES) USING MOLECULAR DESCRIPTORS

B. Souyei[*] and M. Korichi

Laboratoire de Valorisation et Technologie des Ressources Sahariennes.
Département Sciences de la Matière. Faculté des Sciences et de la Technologie
Université d'El-Oued. B.P.789 El-Oued 39000 Algérie

ABSTRACT

A quantitative structure–property relationship (QSPR) study is carried out to develop correlations that relate the molecular structures of organic compounds (Aliphatic Alkanes) to their normal boiling point (NBP) and two correlations were proposed for constitutionals and connectivity indices Models. The correlations are simple in application with good accuracy, which provide an easy, direct and relatively accurate way to calculate NBP. Such calculation gives us a model that gives results in remarkable correlations with the descriptors of blokes constitutionals (CON), and connectivity indices (CI) ($R^2 = 0.950$,   = 0.766) ($R^2 = 0.969$,   = 0.782) respectively.

Keywords: QSPR; Model; Molecular descriptors; Normal Boiling point; multiple linear regression "Stepwise" .

1. INTRODUCTION

Being firstly and easily measured as a property of a compound, the normal boiling point (NBP) is an important property for the simulation of processes in chemical and petroleum industries and used to characterize and identify a new compound, and to estimate other physical properties, such as critical temperatures, vapor pressures, enthalpies of vaporization, etc[1-2].

_____

Author Correspondence, e-mail: belgacem3963_alg@yahoo.fr
ICID: 1078169

To explain relative boiling points we must take into account a number of properties for each substance. The properties include molar mass, structure, polarity and hydrogen bonding ability [3]. It is well-known that NBP is determined by the intermolecular interactions, and by the difference in the molecular internal partition function between gas and liquid phases at the boiling point and also it is proportional to the enthalpy of vaporization and inversely proportional to the entropy of vaporization[7]. Consequently, the NBP depends indirectly on the chemical structure of the molecules . It is not surprising that numerous methods have been developed to estimate the NBP of a compound from its structure [1]. The quantitative structure-property relationship (QSPR) model has been quite extensively reported in the literature to predict the boiling point[4-5]. A key point in the QSPR studies is the accurate and simple characterization of the structural features that are related to the observed property, known as the molecular descriptors [5-6].

In this work we aim to use molecular descriptors as an alternative approach in the prediction of boiling point (NBP) by the mean of linear regression techniques.

2. METHODOLOGY

Our input data set contains 170 molecules of aliphatic alkanes. The geometry of all molecules were optimized by the molecular mechanic MM+ and semi-empirical method AM1 in Hyper chem software version 7.5 and used to calculate the molecular descriptors of the input data set molecules (topological charge indices (TCI), connectivity indices (CI), topological indices (TI), constitutionals (CON), Atom centred fragments (ACF), using the dragon software (TALETE, 2007 version 7.5) .

Figure (1) summarizes the proposed methodology. According to figure 1, the steps of the proposed methodology are:

2.1. Molecular descriptors calculation

- Molecular descriptors of the input data set molecules (topological charge indices (TCI), connectivity indices (CI), topological indices (TI), constitutionals (CON), atom centred fragments (ACF).

- The total number of molecular descriptors is **173**, and is divided between the different descriptors: topological charge indices (**21**), connectivity indices (**33**), topological indices (**78**), constitutionals (**30**) and Atom centred fragments (**11**).

2.2. Complete Correlation Analysis:

Select a subset of linearly independent descriptors. Descriptor dependency is evaluated using the Dragon software (TALETE, 2007 version 5.5) by setting a predefined value R $_{max}$ (In this work, R $_{max}$ = 0.97) below which descriptors are considered linearly independent.

3.3. Multiple Linear Regression:

This analysis is divided into simple regression analysis (training set 70%) and validation of the model by 48 molecules that present 30% of overall set. We have made calculations using the XLSTAT 2010 software and reported the results for these descriptors with reference to the experimental values. The models were calculated using the method step by step selection of variables for multiple linear regression "Stepwise". With the probability of entry is = 0.05

Fig.1.
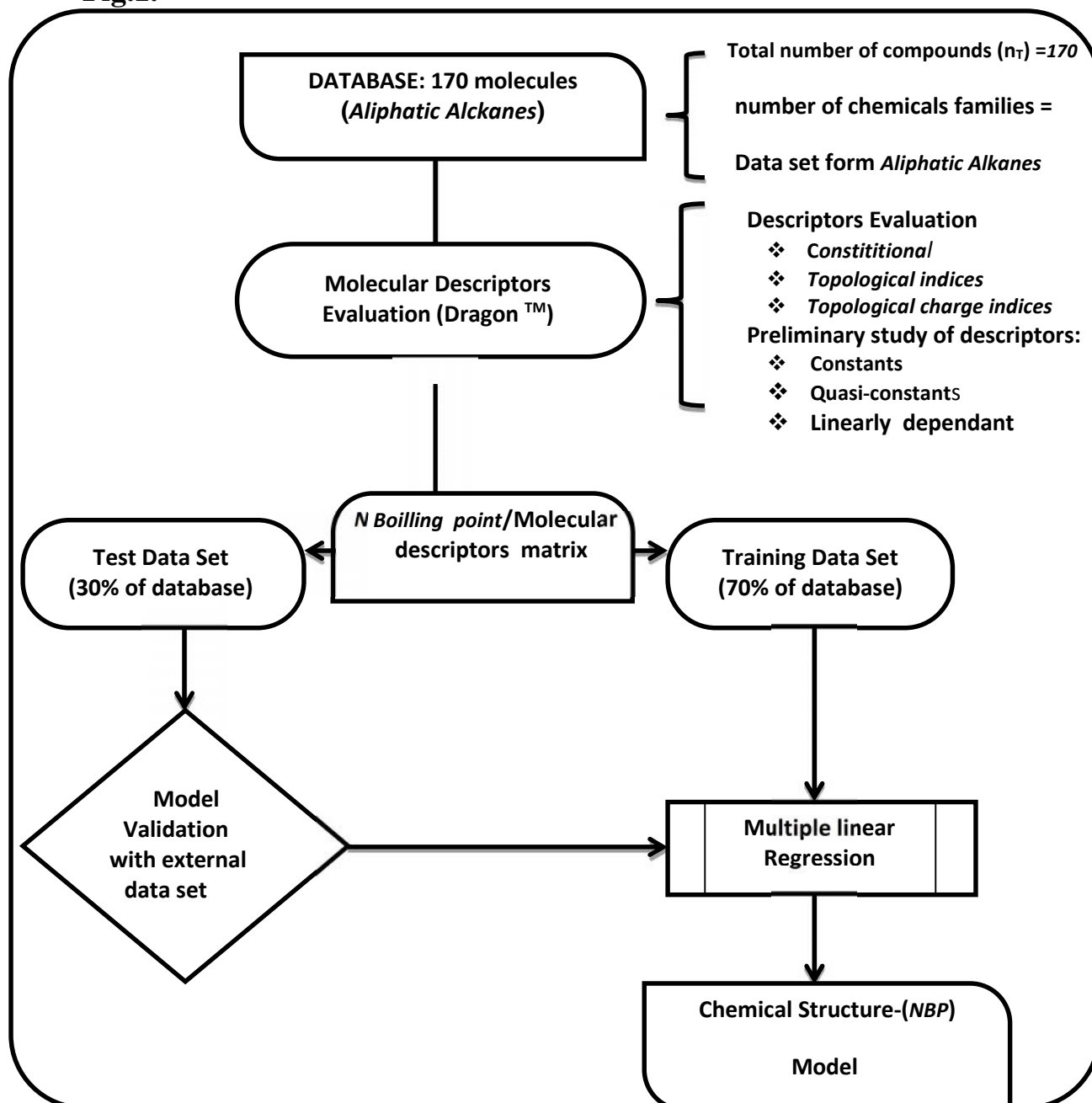
3. RESULTS AND DISCUSSIONS

The comparison between descriptors blocks is based on the correlation coefficient R^2, the number of variables and uncorrelated molecules within each block.

According to the statistical analysis of **BP** [pred], **BP** [exp] and standardized residuals values and the interpretation of graphs (**BP** [pred]/ **BP** [exp]), we can see that the residue values range from (-2 to +2) and values that are outside this range are outliers shows that the molecules not well correlated**.**

These molecules have:

- irregular two-dimensional chemical structures;

- hydrogen bonds, van deer Waals interaction.,

The quality of the prediction model depends on :

-The number of variables (parameters).

- The value of R^2 indicates the degree of correlation property (**BP**) with the structural parameters of different blocks.

After the comparison between the different blocks between, we can say that the descriptors from 2D blocks (constitutionals (CON) and topological indices (TI)), give adequate representation to the molecular structure.

Table 1. Molecular descriptors, correlations coefficients (R^2), average absolute deviation (AAD), standard deviation (STD), number of variables used in the models (NVU) and the Number of no correlated Molecules-out of 95%- (NMNC) [The number of used molecules is 170] .

Descriptors	ACF	Con	TI	TCI	CI
R^2	0832	0.950	0.850	0.761	0.969
AAD	0.553	0.575	0.577	0.653	0.583
STD	0.757	0.766	0.769	0.771	0.782
NVU	6	3	6	6	8
NMNC	7	3	17	10	3
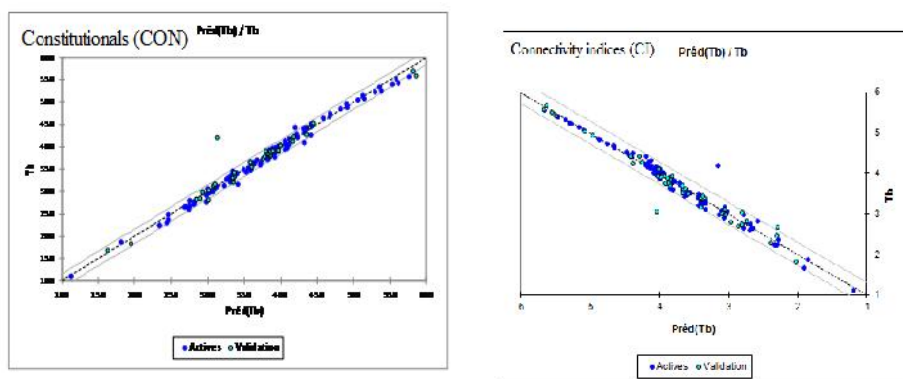
4. CONCLUSION

In this work we present different ways to estimate the **billing point (BP).** We made the correlation of different types of descriptors and contribution to the calculation by QSPR method. This work was performed on 173 different parameters of 5 blocks and a database contains 170 molecules of Aliphatic Alkanes.

For prediction we found a well mathematical relationship is (**BP** = f (structural parameters), **BP** = f(b + sum (**ai** $\times$ **xi**). This model gives results in remarkable correlation with the descriptors of blocks constitutionals (CON) and connectivity indices (CI), (R2 = 0.950,   = 0.766), (R2 = 0.969,   = 0,782) respectively.

5. REFERENCES

[1] Chen Q., Kou Y. W., Panand F., Yuan J. Y., J. Iran. Chem. Soc. 7, 2010, 1012-1020.

[2] Eleni. P. Epaminondas V., Kostis M., Dimitrios T., Fluid Phase Equilibria. 01/2006, 248(1), 70-77.

 [3] Chemistry in the Community, American Chemical Society, second edition, 1995**.**

[4]Maykel P.G., Andrey A. T ., Pablo R. D., Eduardo A. C., Molecules 2004, 9, 1019-1033.

[5] Castro E. A., Toropov A. A., Toropova A. P., Akhmerov R. Z., Kragujevac J. Sci. 31 (2009), 33-43.

[6]Gleeson M. P., Modi S., Bender A., Marchese Robinson R. L., KirchmairJ., Promkakaew M., Hannongbua S., Glen R. C., Curr. Pharm. Des. 2012, 18, 1266–1291.

[7] Michael D. G., Kurt. B., Dane J., Relative Volatility Paper, Performance Products, Taminco 2012.