STUDIA UBB. CHEMIA, LV, 4, 2010

THE RELATIONSHIP BETWEEN ENERGY CALCULATIONS AND BOILING POINTS OF N-ALKANES

LORENTZ JÄNTSCHI[a], SORANA D. BOLBOACĂ[a,b]

ABSTRACT. The relationship between energy calculations and boiling points was studied on a set of fourteen n-alkanes. The correlation analysis clearly showed that the best relationship is not linear. The regression analysis showed that a dose-response logistic function provided a very good agreement between the boiling points of alkanes and their heat of formation.

Keywords: *regression, correlation, alkanes, boiling point, energy*

INTRODUCTION

Boiling point, the temperature at which the vapor pressure of the liquid equals the environmental pressure surrounding the liquid [1], of organic compounds is an important property since it can provide information about other physical properties and structural characteristics [2]. Molecules with strong intermolecular forces are known to have higher boiling points [2].

The boiling point of alkanes, chemical structures with a C_nH_{2n+2} generic formula, increases with the chain length (number of carbon atoms).

The relationship between the boiling points of alkanes and other properties or descriptors have previously been studied using simple or multiple linear regression models [3-5] or non-linear models [6]. Since the boiling point of alkanes is determined by their molecular weight, this property shows a linear relationship with the size of the molecules [7]. Kozioł obtained, on a set of fourteen n-alkanes, a non-linear model with five descriptors having a determination coefficient of 0.9993 [6]. Moreover, simple exponential models estimated the critical temperature, pressure, and volume of alkanes as function of the normal boiling point and molecular weight [8].

The present study is aimed to carry out correlation and regression analyses in order to establish the relationships between the calculated energy and the boiling points of n-alkanes (an "easy to predict" property).

---

[a] *Technical University of Cluj-Napoca, 103-105 Muncii Bvd., RO-400641 Cluj-Napoca, Romania, lori@academicdirect.org*

[b] *"Iuliu Hațieganu" University of Medicine and Pharmacy Cluj-Napoca, 13 Emil Isac, RO-400023 Cluj-Napoca, Romania, sbolboaca@umfcluj.ro*

RESULTS AND DISCUSSION

The results of the correlation analyses are presented in Table 1. The dipole moment property was excluded from further analyses since the Pearson correlation coefficient was of -0.0391. The analysis of the obtained correlation coefficients revealed that Spearman and Gamma correlation coefficients had higher values compared to the Pearson correlation coefficients.

Table 1. Results of correlation analysis

X (Y= boiling point)	r (p)	ρ (p)	Γ (p)
heat-of-formation	0.9515958 ($1.67\cdot10^{-7}$)	1^*	1^*
scf-binding-energy	0.9499073 ($2.05\cdot10^{-7}$)	1^*	1^*
total-energy	0.9498675 ($2.06\cdot10^{-7}$)	1^*	1^*
scf-atom-energy	0.9498641 ($2.06\cdot10^{-7}$)	1^*	1^*
scf-electronic-energy	0.9060543 ($8.09\cdot10^{-6}$)	1^*	1^*
scf-core-energy	0.8992529 ($1.21\cdot10^{-5}$)	1^*	1^*
dipole-moment	-0.0391090 (0.8943)	0.0681 (0.8094)	0.0989 (0.9618)

Correlation coefficients: r = Person; ρ = Spearman; Γ = Gamma
* $p < 10^{-7}$;

The 0.9515958 value of the Pearson correlation coefficient revealed that the linear relationship with the heat of formation was able to explain almost 91% of boiling points variation of the studied n-alkanes, which is a good estimation. Since the Spearman correlation coefficient was equal to the Gamma correlation coefficient and both of them were higher than the Pearson correlation coefficient, the relationship between boiling points and energy calculations could be non-linear.

Non-linear regression analysis was carried out in order to identify the type of relationship between the boiling points of alkanes and energy calculations. The best performing models, in terms of determination coefficients, F-value and coefficient significance proved to be of the *dose-response logistic function* type. The top three models in terms of the above-presented criteria are shown in Table 2.

The analysis of the results in Table 2 revealed that the best performing model, able to explain the boiling points of alkanes (as estimator) used the heat of formation (as predictor, H_F) through a dose-response logistic function. As it can be observed, a four-variable equation was able to fully predict the variation of boiling points as function of the heat of formation. The smallest difference between the determination coefficient and the adjusted determination coefficient was obtained using the first equation (boiling point as function of the heat of formation). The smallest value of the standard error was of 0.33°C and provided by the first equation (boiling point as function of the heat of formation). Note that the highest t-values associated to the coefficients and the smallest values of the standard errors were obtained when the boiling points were investigated as function of the heat of formation.

Table 2. Regression analysis results

Type		r^2	r^2_{adj}	F (FitStErr)	C	Value [95%CI]	StErr	t
Y	X							
DoseRspLgstc $\hat{Y} = a_0+a_1/(1+(x/a_2)^a_3)$								
B_P	H_F	0.999997	0.999996	1090130	a_0	1142.31 [1111.59; 1173.03]	13.78	82.85
				(0.32797)	a_1	-1435.64 [-1470.43; -1400.85]	15.61	-91.94
					a_2	-191.47 [-200.82; -182.11]	4.20	-45.59
					a_3	0.7518 [0.7386; 0.7656]	0.01	121.71
B_P	T_E	0.999864	0.999823	24478	a_0	-324.89 [-367.34; -282.43]	19.06	-17.05
				(2.18849)	a_1	1836.08 [1332.98; 2339.17]	225.80	8.13
					a_2	-179833.96 [-305299; -54369]	56313	-3.19
					a_3	-0.6190 [-0.7225; -0.5155]	0.046	-13.32
B_P	SBE	0.999857	0.999814	23351	a_0	-359.58 [-416.26; -302.91]	25.44	-14.14
				(2.24065)	a_1	1925.18 [1315.38; 2534.99]	273.70	7.03
					a_2	-14657.09 [-26730; -258]	5418.9	-2.70
					a_3	-0.5950 [-0.7137; -0.4764]	0.0532	-11.15

DoseRspLgstc = dose-response logistic function;
B_P = boiling point; H_F = heat-of-formation; T_E = total-energy; SBE = scf-binding-energy;
r^2 = determination coefficient; r^2_{adj} = adjusted determination coefficient; F = F-value;
C = coefficient; 95%CI = 95% coefficient confidence interval; StErr = standard error;
t = t-value

The graphical representation of the best performing model (B_P^= $(1142.31\pm30.72)-(1435.6\pm34.79)/(1+(H\_F/(-191.47\pm9.35))^{(0.7518\pm0.0132)})$) is presented in Figure 1.

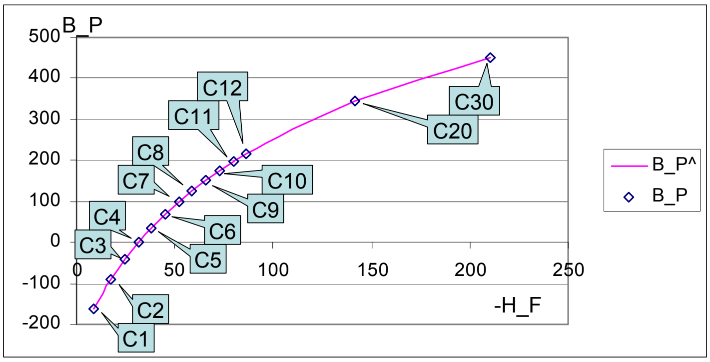

Figure 1. Boiling points of alkanes as heat of formation function

The analysis of Figure 1 revealed that the identified dose-response logistic function is the best one in estimating the relationship between the heat of formation and the boiling points of the studied n-alkanes. This statement is also supported by the value of the correlation coefficient associated to the model (see Table 2). A statistically significant linear relationship could also be identified between boiling points and the heat of formation, but this relationship had lower performances compared to the best scoring dose-response logistic function ($r^2 = 0.9062$, $F = 116$, $p = 1.6 \cdot 10^{-7}$, standard error of estimated = 52.44).

The estimated boiling points when the first equation was used (boiling point as function of the heat of formation), abbreviated as B_P^, and the measured boiling points, abbreviated as B_P, is graphically presented in Figure 2.

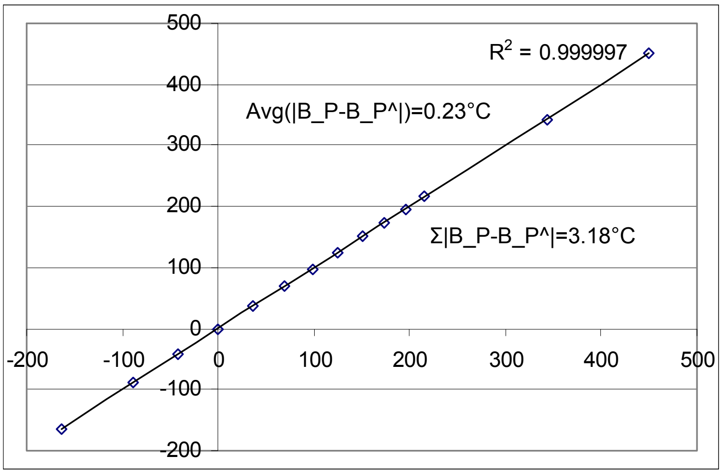

Figure 2. Estimated (horizontal) versus measured (vertical) boiling points using the dose-response logistic function

The validity and reliability of the best performing relationship obtained in the study on n-alkanes is supported by the smallest value of the absolute value of residuals (equal to 0.23°C) and by the sum of the absolute difference of residuals (equal to 3.18°C) (Figure 2). Moreover, the sum of residuals was 0.01°C while the squared sum of residuals was 1.08.

The objective of this research was met as soon as the best model able to estimate the boiling points of alkanes as functional dependence on energy calculations was identified. The value of the Person correlation coefficient, which proved to be smaller in comparison to the Spearman and Gamma correlation coefficients, determined the investigation of non-linear relationships even if the linear relationship was statistically significant. A dose response logistic function proved to better explain the boiling points as function of energy calculations for the studied n-alkanes when the molecules were prepared for analysis by applying the mm+ as molecular mechanics and the AM1 as semi-empirical method.

CONCLUSIONS

If ρ^2(Spearman), Γ^2(Gamma) >r^2(Pearson), the relationship between variables is not linear; non-linear relationships must always be checked. Thus, the best performing relationship between boiling points and the energy calculations of the investigated n-alkanes was expected not to be linear.

A functional dependence was identified between boiling points and the energy calculations of the investigated n-alkanes. This functional dependence proved to be a dose-response logistic function when mm+ molecular mechanics and AM1 semi-empirical methods were used to prepare the studied n-alkanes for analysis.

The following model was identified as the model with the highest performance:

$$\text{B\_P\^{}} = (1142.31\pm30.72)-(1435.6\pm34.79)/(1+(\text{H\_F}/(-191.47\pm9.35))^{(0.7518\pm0.0132)}),$$

where B_P^ is the estimated boiling point and H_F is the heat of formation. The validity of the model is supported by the small value of the standard error, the high F-value and the small p-value.

EXPERIMENTAL SECTION

Fourteen normal alkanes (C_1-C_{12}, C_{20}, C_{30}), chemical compounds consisting of carbon and hydrogen elements, were analyzed (see Table 3).

Table 3. Characteristics of alkanes: boiling point, dipole-moment, total-energy, atom-energy, binding-energy, core-energy, electronic-energy, and heat-of-formation

Name	Formula	B_P	D_M	T_E	SAE	SBE	SCE	SEE	H_F
Methane	CH_4	-164	$1.12\cdot10^{-6}$	-4225	-3837	-388	4619	-8844	-9
Ethane	C_2H_6	-89	$6.87\cdot10^{-7}$	-7821	-7149	-672	13638	-21459	-18
Propane	C_3H_8	-42	$4.28\cdot10^{-3}$	-11415	-10461	-954	26313	-37727	-24
Butane	C_4H_{10}	-0.5	$1.01\cdot10^{-7}$	-15008	-13773	-1236	41607	-56615	-31
Pentane	C_5H_{12}	36	$6.28\cdot10^{-3}$	-18602	-17084	-1518	59034	-77636	-38
Hexane	C_6H_{14}	69	$3.06\cdot10^{-7}$	-22196	-20396	-1800	78191	-100387	-45
Heptane	C_7H_{16}	98	$6.57\cdot10^{-3}$	-25790	-23708	-2082	98835	-124624	-52
Octane	C_8H_{18}	125	$1.52\cdot10^{-7}$	-29383	-27020	-2364	120757	-150141	-59
Nonane	C_9H_{20}	151	$6.65\cdot10^{-3}$	-32977	-30331	-2646	143819	-176796	-66
Decane	$C_{10}H_{22}$	174	$3.95\cdot10^{-7}$	-36571	-33643	-2928	167892	-204463	-73
Undecane	$C_{11}H_{24}$	196	$8.13\cdot10^{-3}$	-40165	-36955	-3210	192888	-233052	-80
Dodecane	$C_{12}H_{26}$	216	$1.35\cdot10^{-7}$	-43758	-40267	-3492	218724	-262482	-86
Eicosane	$C_{20}H_{42}$	343	$8.61\cdot10^{-7}$	-72508	-66760	-5748	449165	-521673	-142
Triacontane	$C_{30}H_{62}$	450	$1.59\cdot10^{-6}$	-108445	-99878	-8567	779447	-887893	-210

B_P = boiling point; D_M = dipole-moment; T_E = total-energy;
SAE = scf-atom-energy; SBE= scf-binding-energy; SCE = scf-core-energy;
SEE = scf-electronic-energy; H_F = heat-of-formation.

Eight properties of the above-mentioned alkanes were investigated: boiling point [°C] [9], total-energy (T_E) [kcal/mol], dipole-moment (D_M) [Debyes], scf-atom-energy (SAE) [kcal/mol], scf-binding-energy (SBE) [kcal/mol], scf-core-energy (SCE) [kcal/mol], scf-electronic-energy (SEE) [kcal/mol], and heat-of-formation (H_F) [kcal/mol]. Except for the boiling points, all the other properties were calculated with HyperChem v. 8.0 using the following criteria: optim-converged=true, molecular mechanics method: mm+ [10], and semi-empirical method: AM1 [11].

Correlation and regression analyses were carried out in order to meet the objective of the study. Pearson ("r") [12], Spearman ("ρ") [13] and Gamma ("Γ") [14] correlation coefficients were used to find the power and the sign of the relationship between boiling points and the investigated properties.

Regression analyses were carried out with the SlideWrite Plus software. The following possibilities of regression search were used:

- *Linear:* ▪ Linear Group; ▪ Exponential Group; ▪ Power Group; ▪ Polynomial Group.
- *Nonlinear:*
    - *Standard*: ▪ User-Defined (any function defined by the user); ▪ Exponential – $Y=a_0+a_1*exp(-x/a_2)$; ▪ Power - $Y=a_0+a_1*x\hat{}a_2$.
    - *Transitional*: ▪ 1-Site Ligant – $Y=a_0*x/(a_1+x)$;
      ▪ Cumulative – $Y=a_0+a_1*0.5*(1+erf((x-a_2)/\sqrt{}(2)*a_3))$;
      ▪ DoseRspLgstc - $Y=a_0+a_1/(1+(x/a_2)\hat{}a_3)$;
      ▪ Photosynthesis - $Y=a_0*a_1*x/(a_0+a_1*x)$;
      ▪ PH Activity – $Y=(a_0+a_1*10\hat{}(x-a_2))/(1+10\hat{}(x-a_2))$;
      ▪ Sigmoidal – $Y=a_0+a_1/(1+exp(-(x-a_2)/a_3))$.
    - *Peak*: ▪ Erfc Peak, Gaussian – $Y=a_0+a_1*exp(-0.5*((x-a_2/a_3)^2)$;
      ▪ Logistic Peak – $Y=a_0+a_1*4*\ (exp(-(x-a_2)/a_3))/(1+exp(-(x-a_2)/a_3))^2$;
      ▪ Log-Normal – $Y=a_0+a_1*exp(-0.5*(ln(x/a_2)/a_3)^2)$;
      ▪ Lorentzian – $Y=a_0+a_1/(1+((x-a_2)/a_3)^2)$.
    - *Waveform*: ▪ SineWave – $Y=a_0+a_1*sin(2*pi*x/a_3+a_2)$;
      SineWaveSquared – $Y=a_0+a_1*(sin(2*pi*x/a_3+a_2))^2$
- *User-Defined:* allows to define any equation with a maximum of 7 coefficients.

ACKNOWLEDGMENTS

Financial support is gratefully acknowledged to CNCSIS-UEFISCSU Romania (project PNII-IDEI1051/202/2007).

REFERENCES

1. D.E. Goldberg, 3,000 Solved Problems in Chemistry (1st ed.). McGraw-Hill. ISBN 0-07-023684-4. Section 17.43, 1988, pp. 321.
2. M.R. Riazi, T.E. Daubert, *Hydrocarbon Processing,* **1980**, *59(3)*, 115.
3. E.S. Souza, C.A. Kuhnen, B.S. Junkes, R.A. Yunes, V.E.F. Heinzen, *Journal of Chemometrics,* **2010**, *24(3-4)*, 149.
4. D. Ciubotariu, M. Medeleanu, V. Vlaia, T. Olariu, C. Ciubotariu, D. Dragos, S. Corina, *Molecules,* **2004**, *9(12)*, 1053.
5. S.D. Bolboacă, L. Jäntschi, *International Journal of Pure and Applied Mathematics,* **2008**, *47(1)*, 23.
6. J. Kozioł, *Polish Journal of Chemistry,* **2009**, *83(12)*, 2173.
7. R.T. Morrison, R.N. Boyd. Organic Chemistry, 6th ed. (New Jersey: Prentice Hall, 1992.
8. F. Vejahati, M.B. Nikoo, S. Mokhatab, B.F. Towler, *Petroleum Science and Technology,* **2007**, 25(9), 1115.
9. C.E. Ophardt. *Virtual Chembook*, Elmhurst College, Elmhurst, IL, USA, 2003.
10. A. Hocquet, M. Langgard, *Journal of Molecular Modeling,* **1998**, *4*, 94.
11. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, *Journal of the American Chemical Society,* **1985**, *107*, 3902.
12. K. Pearson, *Philosophical Magazine,* **1900**, *50*, 157.
13. C. Spearman, *American Journal of Psychology*, **1904**, *15*, 201.
14. L.A. Goodman, W.H. Kruskal, *Journal of the American Chemical Society,* **1963**, *58*, 310.