Research Article

QSPR Model for Predicting Flash Point of Some Organic Hydrocarbons

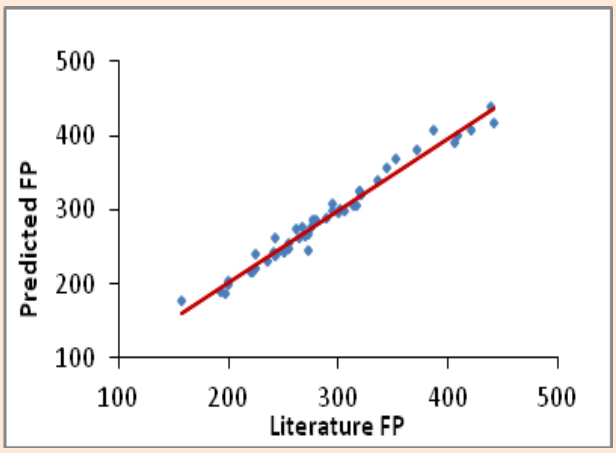
Elshafie A. M. Gad*, Jabir H. Al-Fahemi, and Nahlah A. Albis

Chemistry Dept. Faculty of Applied Science, Umm-Al-Qura University, Mekkah, Saudi Arabia

Abstract

Quantitative structure property relationship, QSPR is performed as a means to predict flash point of hydrocarbons via correlating this property to parameters calculated from molecular structure. Flash points of different hydrocarbons are compiled from literature survey. The chemical structure of hydrocarbons are geometry optimized using HyperChem; molecular modelling system for windows Version 8.0.4. Molecular descriptors are calculated such as polarizability α, total connectivity T.C, molar volume V, molar refractivity R, Molecular Mass M, wiener index W, partition coefficient and partition coefficients log p. Principal component analysis PCA and multiple linear regression technique MLR were performed to examine the relationship between the selected descriptors and the flash point of hydrocarbons. The results of PCA explain the inter-relationships between flash point and different variables. The linear relationship between the selected descriptors and flash point was modelled according to the better statistical results. The best model has coefficient of determination (R^2 = 0.975) with statistical significance (F = 338.4) The obtained a QSPR model allows estimating of flash point for unknown hydrocarbons using theoretical-calculated descriptors.

Keywords: flash point, principal component analysis, multiple linear regression analysis, quantitative structure property relationship (QSPR), Chemometric data analysis, computational chemistry.



***Correspondence**

Author: Elshafie A. M. Gad, Egyptian Petroleum Research Institute, (EPRI) Nasr City, Cairo. Egypt, 1 Ahmed El-Zomor St.

Email: eamgad_99@hotmail.com

Introduction

The term "flash point" is used to determine the lowest temperature at which a volatile substance can become vaporised into a flammable gas. It is one of the most widely used, important characteristics of the flammability properties of liquids and low-melting substances. It provides a simple, convenient index of the flammability and combustibility of substances and is of importance, since it gives the knowledge needed for the handling and transporting of the compound in bulk quantities.

There are various methods of measuring a flash point, which can be divided into two main categories: open and closed cup flash points. The open-cup method gives a somewhat arbitrary value because of the unpredictable rate of mass transfer between the liquid and the surrounding atmosphere, but still provides the best match to reality. The closed-cup method produces the most consistent results, because the FP depends only on the vapour pressure and the heat effect of the initial oxidation.

Quantitative structure property relationship (QSPR) and topological indices have been used to predict flash point properties of different classes of solvents by Patel, S. J. et al. [1-2]. Multiple linear regression and back-propagation

neural network analysis were used to model the flash point. The neural network model showed higher accuracy (training set, r = 0.948, R-2 = 0.898). However, there are certain limitations associated with using QSPR in CAMD which have been discussed and need further work.

The flash point of alkanes was modelled [3] by a set of molecular connectivity indices, modified molecular connectivity indices and valance molecular connectivity. A stepwise multiple linear regression method was used to select the best indices. The predicted flash points are in good agreement with the experimental data, with the average absolute deviation 4.3 K.

Rahimi et al. [4] have attempted to develop a simple and fast multiple linear regression model. The molecular descriptors, which cover different information of molecular structures, were calculated by Dragon software. Katritzky et al. [5] studied a QSPR of the flash point using experimental boiling point. The reported relationship gives a three-parameter flash point equation with a R^2 value of 0.9247. Katritzky et al. [6] developed QSPR models for the flash points using geometrical, topological, quantum mechanical and electronic descriptors calculated by CODESSA PRO software. Paralikas et al. [7] underlined that there is no single property to describe or appraise flammability and fire risk of materials. Flash points of various classes of organic compounds were studied using fragmental approach in the framework of QSPR methodology. The fragmental descriptor based regression and neural network models for flash point prediction are proposed [8].

A QSPR model (R2 = 0:9669 and s = 12:691) for the prediction of flash points is developed [9]. Genetic Algorithm-based Multivariate Linear Regression (GA-MLR) technique is used to select four chemical structure-based molecular descriptors from a pool containing 1664 molecular descriptors. Keshavarz et al. [10] proposed model to be used for different hydrocarbons including cyclic and acyclic compounds with complex molecular structures. A QSPR model was developed to predict the flash points of organic compounds containing a collection of 57 functional groups were selected as the molecular descriptors [11].

In this study, The calculated molecular descriptors such as polarizability α, total connectivity T.C, molar volume V, molar refractivity R, Molecular Mass M, wiener index W, partition coefficient and partition coefficients log p were investigated to find out the optimum model for flash point prediction.

Methodology and Data processing

Experimentally determined flash point values of the selected hydrocarbons were quoted from literatures [1-4] (**Table 1**). All calculation in this study were performed using HyperChem; molecular modelling system for windows Version 8.0.4. The geometry of all hydrocarbons were optimized using AM1 semi-empirical method. The calculated molecular descriptors are shown in Table 1, namely Polarizability α, total Connectivity T_C, molar volume V_m, molar Refractivity R_m, molecular mass M, Wiener index W, octanol/water distribution coefficient log p. The relative importance of the descriptors can be confirmed by looking at the correlation coefficients matrix. The higher the correlation coefficient is significant values regardless its sign positive or negative. From the obtained correlation coefficients matrix shown in **Table 2**, it is quite clear that the selected descriptors have good correlation with the flash point.

Table 1 Polarizability α, total Connectivity T.C, volume V, Refractivity R, Mass M, wiener index W, partition coefficient log p, and estimated values of Flash point FP of hydrocarbons

no	Hydrocarbons	α	T.C	V	R	M	W	Log p	FP(K)
1	Pentane	9.95	0.353	375.8	24.81	72.15	20	2.49	224.1
2	Hexane	11.7	0.25	429.2	29.41	86.18	35	2.88	250.1
3	Heptane	13.6	0.176	482.9	34.01	100.2	56	3.28	272.1
4	Octane	15.4	0.125	536.3	38.61	114.2	84	3.67	289.1
5	Nonane	17.2	0.088	589.99	43.21	128.2	120	4.07	314.1
6	Decane	19.1	0.062	643.41	47.81	142.2	165	4.47	319.1
7	Dodecane	22.7	0.031	750.48	57.01	170.3	286	5.26	344.1
8	Cyclohexane	11.0	0.125	373.5	27.61	84.16	27	2.38	255.1

9	Cyclohexene	10.8	0.125	361.86	28.72	82.15	27	2.12	243.1
10	Benzene	10.4	0.125	331.29	30.96	78.11	27	1.6	262.1
11	Methylbenzene	12.2	0.102	383.6	35.24	92.14	42	1.75	280.1
12	1,2-dimethylbenzene	14.1	0.083	426.9	39.52	106.1	60	1.9	305.1
13	1,4-dimethylbenzene	14.1	0.083	435.64	39.52	106.1	62	1.9	300.1
14	1,3,5-trimethylbenzene	15.9	0.068	488	43.8	120.1	84	2.06	317.1
15	Isopropylbenzene	15.9	0.0589	473.18	44.39	120.1	88	2.48	319.1
16	Tridecane	24.6	0.022	804.14	61.62	184.3	364	5.66	352.1
17	Tetradecane	26.4	0.0156	857.56	66.22	198.3	455	6.05	372.1
18	Hexadecane	30.13	0.007	964.63	75.42	226.4	680	6.84	408.1
19	Heptadecane	31.97	0.005	1018.2	80.02	240.4	816	7.24	421.1
20	Nonadecane	37.47	0.001	1178.7	93.82	282.5	1330	8.43	441.1
21	Isopentane	9.95	0.408	362.82	24.75	72.15	18	2.42	221.1
22	Isohexane	11.78	0.288	416.48	29.36	86.18	32	2.82	250.1
23	3-methylpentane	11.78	0.288	409.33	29.36	86.18	31	2.82	241.1
24	2,3-dimethylbutane	11.78	0.3333	403.44	29.3	86.18	29	2.75	244.1
25	Buta-1,3-diene	7.73	0.5	290.85	20.29	54.09	10	1.66	197.0
26	But-2-ene	7.92	0.5	309.42	21.32	56.11	10	1.83	199.8
27	(z)-but-2-ene	7.92	0.5	256.06	21.32	56.11	10	1.83	200.1
28	But-1-ene	7.92	0.5	309.27	20.25	56.11	10	1.87	193.1
29	2-methylprop-1-ene	7.92	0.577	306.34	19.93	56.11	9	1.63	157.1
30	(z)pent-2-ene	9.76	0.3535	156.09	25.92	70.13	20	2.23	255.1
31	Pent-1-ene	9.76	0.353	362.92	24.85	70.13	20	2.27	222.15
32	Dec-1-ene	18.93	0.0625	631.94	47.86	140.2	165	4.25	320.1
33	Hept-1-ene	13.43	0.176	470.47	34.05	98.19	56	3.06	264.1
34	Cyclooctane	14.68	0.0625	446.68	36.81	112.2	64	3.17	301.1
35	Cyclopenta-1,3-diene	8.79	0.176	301.91	25.24	66.1	15	1.46	273.1
36	Cyclopentane	9.18	0.176	334.15	23.01	70.13	15	1.98	236.1
37	Cyclopentene	8.98	0.176	318.57	24.12	68.12	15	1.72	243.1
38	2-methylheptane	15.45	0.144	523.56	38.56	114.2	79	3.61	277.1
39	4-vinylcyclohexene	14.3	0.072	440.75	37.92	108.1	64	2.63	294.1
40	Ethylbenzene	14.1	0.072	432.51	39.84	106.1	64	2.15	295.1
41	Undecane	20.96	0.044	697.07	52.41	156.1	220	4.86	335.1
42	Pentadecane	28.3	0.011	911.22	70.82	212.4	560	6.45	405.1
43	Octadecane	33.3	0.003	1071.7	84.62	254.5	969	7.64	439.1
44	Tricosane	42.98	0.0006	1339.5	107.6	324.6	2024	9.62	386.1
45	2,2-dimethylbutane	11.78	0.353	400.03	29.23	86.18	28	2.85	225.15
46	2-methylhexane	13.62	0.204	469.9	33.96	100.2	52	3.21	270.1
47	2,4,4-trimethylpentene	15.26	0.204	472.45	38.16	112.2	66	3.19	267.1
48	2,3,4-trimethylpent-2ene	15.26	0.192	477.43	39.04	112.2	65	2.86	275.1

Table 2 Correlation coefficients matrix of flash point and selected descriptors

	α	T.C	V	R	M	W	log P	FP
α	1							
T.C	-0.694	1						
V	0.990	-0.666	1					
R	0.997	-0.716	0.982	1				
M	0.999	-0.705	0.989	0.997	1			
W	0.920	-0.481	0.910	0.918	0.920	1		
log P	0.973	-0.593	0.979	0.954	0.969	0.902	1	
FP	0.920	-0.858	0.896	0.932	0.920	0.743	0.857	1

Results and discussion

Principal Component analysis:

The biplot shown in **Figure 1** is a visualization technique to investigate the interrelationships between the flash point and descriptors in multivariate data. The placing of a parameter on the biplot shows that flash point is influenced by the vectors that lie near it or in the same side. However those vectors lie perpendicular to flash point have low correlation values. The variables are represented by vectors superimposed on the same plot. The biplot reveals that the parameters such as Polarizability α, volume V, Refractivity R, Mass M, Wiener index W, partition coefficient log p lie close to each other. These parameters lie nearly on the same direction of flash point. It means that these parameters are significantly +ve correlated to flash point. The parameter total connectivity lies nearly on the opposite direction of flash point. It means that the parameter is significantly -ve correlated to flash point. The obtained correlation matrix as shown in Table 2 reveals that T.C gives negative correlation value however, the rest of the descriptors give +ve correlation.

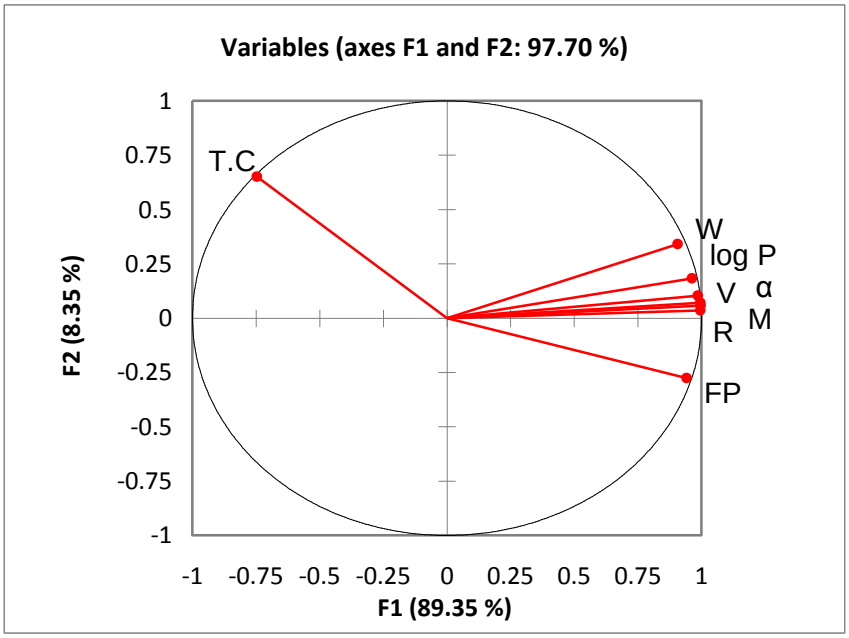


Figure 1 PCA biplot for the selected descriptors against flash point

- *Choice the prober descriptors:* The relative importance of the descriptors can be confirmed by looking at the correlation coefficients matrix. The higher the correlation coefficient is significant values regardless its sign positive or negative. From the obtained correlation coefficients matrix shown in Table 2, it is quite clear that the selected descriptors has good correlation with the flash point.
- *Multiple linear regression analysis MLR:* MLR analysis for the calculated descriptors and flash point values were carried out. The resulting correlation model for prediction the physical property of interest is in the form of the following equation:

$$\text{Property} = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \ldots..$$

$$\boldsymbol{Property} = \beta_0 + \sum_{i=0}^{n} \beta_i \boldsymbol{D_i}$$

$\boldsymbol{\beta_i}$ are the regression Coefficients
$\boldsymbol{D_i}$ are the descriptors

The standard error $\mathbf{s^2}$ expresses the variation of the residuals or the variation about the regression line. Thus the standard error measures the model error. The lower the standard error is the better model .It is observed in **Table 3** that as the number of descriptor combination increase, the standard error decrease. Coefficient of determination R^2 measures the explanatory power of the regression equation. It falls in the range of 0 to 1, where 0 mean the regression accounts for none of the variation and 1 means the relationship was deterministic and the regression accounts for all of the variation. Table 3 shows that R^2 values increases as the number of combined descriptors increases.

Table 3 Standard error s^2, coefficient of determinations R^2 and significance F of different linear regression models

Models	Descriptors combinations	s^2	R^2	F
Model 1	α	26.2	0.85	256.8
Model 2	α + T.C	16.5	0.94	358.6
Model 3	α + T.C + V	16.3	0.943	246.4
Model 4	α + T.C + V + R	16.4	0.944	184.2
Model 5	α + T.C + V + R + M	16.4	0.945	145.2
Model 6	α + T.C + V + R + M + W	12.4	0.969	217.2
Model 7	α + T.C + V + R + M + W+ log p	11.2	0.975	228.4

The t-test measures the statistical significance of the regression coefficients .The higher t-test values correspond to the relatively more significant regression coefficients. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the model. High values of the F-test indicate that the model is statistically significant. The best correlation model was chosen on the basis of the lowest standard error and the highest correlation coefficients and the highest statistical significance. So, the linear relationship between the selected parameters and the flash point was modelled according to the better results. The best regressed model (model 7) has highest coefficient of determination (R^2 = 0.975), highest statistical significance (F= 228.4), and lowest standard errors (s^2 = 11.2).

The obtained a QSPR model allows estimating of flash point for hydrocarbons using theoretical-calculated descriptors. Regression statistic and ANOVA table for the best model (model 7) are represented in **Tables 4 and 5**

respectively. Additionally, **Figure 2** shows graphical representation for the linear relationship between literature values of flash point and the predicted values gained by applying the equation of the model 7.

Table 4 Regression Statistics

Multiple R	0.988
R Square	0.976
Adjusted R Square	0.971
Standard Error	11.289
Observations	48

Table 5 ANOVA

	Df	***SS***	***MS***	***F***	***Significance F***
Regression	7.00	203837.80	29119.69	228.47	3.49902E-30
Residual	40.00	5098.10	127.45		
Total	47.00	208935.90			

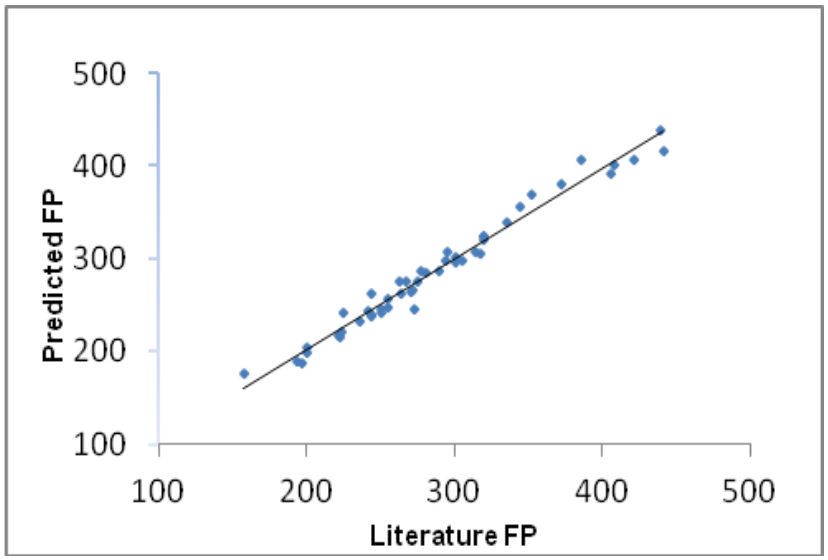


Figure 2 Linear relationship between the literature and predicted values of the flash point

Table 6 shows Regression coefficients, standard errors, t-test statistic and confidence intervals of the best linear regression models. It includes seven descriptors particularly, Polarizability α, total Connectivity T.C, volume V, Refractivity R, Mass M, Wiener index W, partition coefficient log P. Positive values in the regression coefficient indicate that the descriptors contribute positively to the value of flash point FP. Whereas negative values indicate that

the greater the value of the descriptor the lower the value of flash point. The resulting model for prediction of the flash point FP can be written in the following equation:

$$FP = 141.3 - 52.79\ \alpha - 44.47\ T.C - 0.10\ V + 10.16\ R + 4.72\ M - 0.11\ W + 27.55\ \log P$$

Table 6 Regression coefficients, standard errors, t-test statistic and confidence intervals of the best linear regression models. (R^2= 0.975, F= 228.4, s^2= 11.2)

	βi	Coefficients	Standard Error	t -Stat	P-value
Intercept	0	141.37	16.60	8.51	1.61E-10
α	1	-52.79	14.96	-3.53	1.06E-03
T.C	2	-44.47	28.42	-1.56	1.26E-01
V	3	-0.10	0.05	-1.91	6.34E-02
R	4	10.17	2.53	4.02	2.51E-04
M	5	4.72	1.75	2.69	1.03E-02
W	6	-0.11	0.02	-6.50	9.45E-08
log P	7	27.56	8.72	3.16	3.00E-03

An increase in the molecular refractivity, molar mass, and partition coefficient log P, flash point increases. However increasing in the values of other descriptors namely, polarizability, total connectivity, volume, Wiener index, decreases the value of the flash point. According to the best mathematical model, the plot of literature value of flash point estimated values shows a linear correlation.

Eventually, the high correlation coefficient and low standard error for the empirical relationship are quite satisfactory for predicting the flash point based on the polarizability, the total Connectivity, molar volume, the molecular refractivity, the molecular mass, the Wiener index, the partition coefficient log P of hydrocarbons.

Conclusion

- Parameters; namely Polarizability α, volume V, Refractivity R, Mass M, Wiener index W, partition coefficient log p are significantly +ve correlated to flash point. However total connectivity is significantly -ve correlated to flash point.
- The best regressed model (model 7) has highest coefficient of determination (R^2 = 0.975) and highest statistical significance (F= 228.4) with correction factor ±11.3 K
- The empirical QSPR relationship are quite satisfactory for predicting the flash point based on the polarizability, the total Connectivity, the volume, the molecular refractivity, the mass, the wiener index, the partition coefficient log p, of hydrocarbons according to the following equation:

$$FP = 141.3 - 52.79\ \alpha - 44.47\ T.C - 0.10\ V + 10.16\ R + 4.72\ M - 0.11\ W + 27.55\ \log P$$

References

[1] Patel S J, Ng D, Mannan M S, Ind Eng Chem Res 2009, 48, 7378–7387.
[2] Patel S J, Ng D, Mannan M S, Ind Eng Chem Res 2010, 49, 8282–8287.
[3] Atabati M, Emamalizadeh R, Chinese J Chem Eng. 2013, 21, 420–426.
[4] Rahimi M, Nekoei M, Anal Chem Lett 2013, 3, 278-286.
[5] Katritzky A R, Petrukhin R, Jain R, Karelson M, J Chem Inf Comput Sci 2001, 41, 1521–1530.

[6] Katritzky A R, Stoyanova-Slavova I B, Dobchev D A, Karelson M, J Mol Graph Model 2007, 26, 529–536.
[7] Paralikas A N, Lygeros A I, Trans IChemE Part B 2005, 83, 122–134.
[8] Zhokhova N I, Baskin I I, Palyulin V A, Zefirov A N, Zefirov N S, Russ Chem Bull 2003, 52, 1885-1892.
[9] Gharagheizia F, Fareghi Alamdarib R, QSAR Comb Sci 2008, 27, 679–683.
[10] Keshavarz M H, Motamedoshariati H, Ghanbarzadeh M, Chemistry 2011, 20, 58–75.
[11] Pan Y, Jiang J, Wang R, Cao H, Zhao J, QSAR Comb Sci 2008, 27, 1013–1019.

Publication History
Received 07th May 2015
Revised 12th May 2015
Accepted 13th May 2015
Online 30th May 2015