Full Length Research Paper

Modeling and theoretical calculation of liquid heat capacity of alcohols and aldehydes using QSPR

F. Ashrafi, R. Saadati, A. Behboodi Amlashi

Payame Noor University (P.N.U), Farhang Ave., 15 Khordad St., Sari, Iran.

The most important domain part of chemometrics is the studies which relate the properties of molecules to their structural characteristics. In this work we use the quantitative structure – property relationship (QSPR) for providing a linear model for predicting the heat capacity of alcohols and aldehydes in liquid phase at 298 K. To perform this research, a set of 32 alcohol and aldehyde as data series was selected then topologic, electronic and geometric descriptors for data series was calculated. Finally, multiple linear regression method was used for selecting more important descriptors and obtaining convenient model.

Key words: Quantitative structure – property relationship, heat capacity, descriptor, multiple linear regressions, cross validation test.

INTRODUCTION

Quantitative structure – property relationship (QSPR) study is an important section in computational chemistry and uses frequently for predicting physico - chemical and biological activity of organic compounds. To establish the relation between structural characteristics of molecule and its properties the mathematical methods can be used. Multiple linear regression (MLR) is one of the mathematical methods which have an extent application. This method is useful when there is not any interaction between descriptors and their relation with linear defined activity. Heat capacities are applied in reactions for modi-fication of reactants evaluation. In addition, they are use-ful for heat - energy balance design calculation. On the other hand, the tests for determining the heat capacity are expensive and expense much more time. Therefore, we need the models to predict the heat capacity and other physico – chemical properties of molecules. (Ivanova et al., 2004 and Roy et al. 2006), (Gakh et al., 1994) give a model for evaluation the heat capacity of alkanes by using artificial neural network (ANN), which its root mean square error equals to 4.04. (Liu et al., 2000) proposed another model for evaluation of heat capacity of alkanes with the root mean square error of 3.81. Both

of these models were proposed only for alkanes but (Yao et al., 2003) proposed a general model for evaluation of heat capacity of all organic compounds in liquid phase, nevertheless its root mean square error was 17.141. This non linear model can be used for the prediction of liquid heat capacity of Alcohols and Aldehydes. But statistical parameter and prediction ability of present work is better than it.

EXPERIMENTAL METHODS

In parametric methods, one series of digital variables named descriptors were used for evaluation of molecules properties. By using a method either multiple linear regression (MLR) or a non linear method such as artificial neural network (ANN) an equation forms, which states relation between molecule structure and relative variable.

Data series selection

Experimental heat capacities of 32 molecules were selected from existent reference (CRC, 1997) which is shown in Table 1.

Descriptor generation

For modeling, the descriptors which have relation with considered property have to be selected. In the following study 23 topologic

---

*Corresponding author. E-mail: Ferydoun_ashrafi@hotmail.com

Table 1. Data series compounds and the amount of its experimental and MLR calculated values of C_p.

No	Compound	Exp. C_p	Calc. C_p	No	Compound	Exp. C_p	Calc. C_p
		J. Mol^{-1}. K^{-1}				J. Mol^{-1}. K^{-1}	
1	Methanol	81.1	99.36	17	3-Pentanol	239.7	226.51
2	Acetaldehyde	89.0	95.77	18	2-Methyl-2-butanol	247.1	245.45
3	Ethanol	112.3	110.24	19	1-Hexanol	240.4	241.63
4	Ethylene glycol	148.6	149.24	20	Furfural	163.2	157.04
5	Allyl alcohol	138.9	134.58	21	2-Methyl-1-pentanol	248.0	253.81
6	1- Propanol	143.9	142.65	22	4-Methyl-2-pentanol	273.0	260.31
7	2- Propanol	156.5	160.92	23	Benzaldehyde	172.0	196.64
8	1, 2- Propylene glycol	190.8	177.57	24	Metacresol	224.9	222.81
9	Butanal	163.7	161.45	25	Salicylaldehyde	222.0	196.9
10	1-Butanal	177.2	175.13	26	Benzyl alcohol	217.9	223.73
11	2-Butanal	196.9	193.76	27	1-heptanol	272.1	274.58
12	2-methyl-1-Propanol	181.5	194.74	28	1-Octanol	305.2	308.72
13	2-methyl-2-Propanol	218.6	234.53	29	2-Octanol	330.1	329.07
14	Diethylene glycol	244.8	219.44	30	2-Ethyl-1-hexanol	317.5	319.79
15	Cyclopentanol	184.1	198.06	31	1-Decanol	370.6	379.58
16	1-Pentanol	208.1	206.84	32	1-Dodecanol	438.1	447.94

Table 2. Calculated topological descriptors.

Descriptor	Symbol
Wiener index	W
Randic's molecular connectivity index	X
Path one connectivity index	^1x_p
Path two connectivity index	^2x_p
Path three connectivity index	^3x_p
Path four connectivity index	^4x_p
Cluster three connectivity index	^3x_c
Path - cluster four connectivity index	$^4x_{p-c}$
Path one valence index	$^1x_p{}^v$
Path two valence index	$^2x_p{}^v$
Path three valence index	$^3x_p{}^v$
path four valence index	$^4x_p{}^v$
Cluster three valence index	$^3x_c{}^v$
Path - cluster four valence index	$^4x_{p-c}{}^v$
Balaban's J index	J
first order Kappa index	^1k
Second order Kappa index	^2k
third order Kappa index	^3k
Kier flexibility index	Φ
Mean information content index	k_{ic}
Structural information content index	k_{sic}
Complementary information content index	k_{cic}
Bonding information content index	k_{bic}

They were calculated using equations obtained from papers in the different literatures (Wiener, 1947; Randic, 1975; Kier and Hall, 1981; Kier, 1985; Kier, 1980; Balaban, 1981; Basac et al., 1984)

Electronic descriptors that are present in Table 3 provide information about intermolecular forces and interactions. All molecular descriptors were down into Hyperchem (1998) and optimized using MM+ molecular mechanics force field. The resulted geometry was then transferred into Mopac software package (1995) and using keywords such as AM_1, Polar and bonds, Electronic descriptors were obtained.

Geometric descriptors define the molecule dimension and its geometrical shape as exist in Table 4. They were calculated using optimized Cartesian coordinate and Van Der Waals radius of each atom in the molecule by assigned algorithms (Stouch and Jurs, 1986; Rohrbaugh and Jurs, 1987) and programs that written in FORTRAN F77 language.

Feature selection

Some of the 49 descriptors generated for each compound were highly correlated. It was therefore desirable to test each descriptor and eliminate those which show high correlation coefficient (R>0.90). Descriptors that show high correlation were removed from the consideration.

Regression analysis

By using backward multiple linear regression routine implemented in software package SPSS (1992), the linear model have developed, which takes the form:

$$Y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (1)$$

In this equation, Y is the property, that is, the dependent variable, x_1 to x_n represent the specific descriptor, while b_1 to b_n represent the coefficient of those descriptor; b_0 is the intercept of this equation. 5 shows the best MLR model with respect to the number of molecule in data series. The values of the descriptors that were

descriptors, 17 electronic descriptors and 9 geometric descriptors have been calculated.

In Table 2 topological descriptors which provide information about the shape and the number of branches in molecule could be seen.

Table 3. Electronic descriptors calculated by Mopac software.

Descriptor	Symbol
Total energy	ET
Total electronic energy of the molecule	TEE
Core- core repulsion	CORE
Ionization potential	IP
Molecular weight	MW
highest occupied molecular orbital	HOMO
lowest unoccupied molecular orbital	LUMO
Partial charges on the most positive atom	PCPA
Partial charges on the most negative atom	PCAN
Partial charges on the most positive carbon	PCPC
Partial charges on the most negative carbon	PCNC
Electronic density on the most positive atom	EDPA
Electronic density on the most negative atom	EDNA
Dipole moment	DP
Principal moment of inertia	A
Principal moment of inertia	B
Principal moment of inertia	C

Table 4. Geometric descriptors calculated.

Descriptor	Symbol
molecular volume	V_x
Molecular surface area	MSA
Molecular shadow area in the x - y plane	S_{xy}
Molecular shadow area in the x - z plane	S_{xz}
Molecular shadow area in the y - z plane	S_{yz}
Molecular standard shadow area in the x - y plane	SS_{xy}
Molecular standard shadow area in the x - z plane	SS_{xz}
Molecular standard shadow area in the y - z plane	SS_{yz}
Shape factor	$Shape$

Table 5. Specification of the linear model for prediction of liquid heat capacity at constant pressure for alcohols and aldehydes.

Descriptor	Symbol	Mean effect	Coefficient	Std. Error	T value
molecular volume	V	1.01	2.099	.064	32.567
Molecular standard shadow area in the y - z plane	SS_{yz}	- 0.15	- 41.217	15.541	-2.652
cluster three connectivity index	^3X_c	0.026	20.232	4.383	4.616
Partial charge on the most negative carbon	CNC	- 0.09	100.751	30.079	3.350
Path four connectivity index	^4X_p	- 0.03	- 11.667	4.483	-2.630
Intercept	$constant$	----	52.891	15.134	3.495

used in this work were shown in Table 6.

RESULTS AND DISCUSSION

Cross validation test were used for evaluation of model. Statistical parameters of model given in Table 7 and the results were written in Table 8. The overall performance of MLR is evaluated in terms of root mean squared error (RMS) according to the following equation:

$$RMS = \sqrt{\frac{\sum_{i=1}^{ns}\left(Y_i - \hat{Y}_i\right)^2}{n_s}} \quad (2)$$

Table 6. The values of the descriptors that were used in this work.

No [a]	V_x	SS_{yz}	CNC	^4x_p	^3x_c	No [a]	Vx	SS_{yz}	CNC	^4x_p	^3x_c
1	37.094	0.712	-0.0739	0	0	17	104.766	0.639	-0.2157	0.288	0.288
2	47.766	0.735	-0.2889	0	0	18	104.109	0.697	-0.2417	0	1.393
3	54.047	0.877	-0.2152	0	0	19	121.203	0.874	-0.2106	0.676	0
4	62.5	0.880	-0.024	0	0	20	85.203	1.127	-0.2182	0.841	0.204
5	65.172	0.755	-0.2262	0	0	21	121.406	0.706	-0.2114	0.697	0.288
6	70.75	0.888	-0.2117	0	0	22	120.609	0.707	-0.2144	0.943	0.816
7	71.078	0.644	-0.2499	0	0.577	23	102.703	1.123	-0.177	1.595	0.204
8	78.891	0.644	-0.2131	0	0.408	24	108.797	0.717	-0.2113	1.737	0.577
9	81.531	0.847	-0.2338	0.353	0	25	109.953	1.026	-0.2713	1.597	0.402
10	87.484	0.877	-0.2104	0.353	0	26	108.969	0.721	-0.1332	1.595	0.204
11	87.516	0.641	-0.2501	0	0.408	27	138.187	0.893	-0.2108	0.853	0
12	87.234	0.699	-0.211	0	0.408	28	154.953	0.877	-0.2107	1.03	0
13	87.625	0.678	-0.2407	0	2	29	154.828	0.655	-0.2106	0.979	0.408
14	104.906	0.865	-0.0226	0.676	0	30	154.687	0.714	-0.2105	1.189	0.204
15	94.219	0.789	-0.1919	0.552	0.288	31	188.469	0.875	-0.2104	1.384	0
16	104.062	0.890	-0.2114	0.5	0	32	222.344	0.874	-0.2107	1.737	0

[a] the numbers refer to the numbers of the molecules given in table 1.

Table 7. Statistical parameters of model calculated with SPSS software.

R	F-test (regression)	Sig.F	DF Regression
0.994	428.970	.000	5

Table 8. Statistical parameters obtained using cross validation test

R	F-test (regression)	Sig.F	DF Regression	Cross validation test
38.819	0.026	1506.909	0.99	

Where n_s is the number of samples used in creating the QSPR model, y_i is the experimental value and $\hat{y}_i$ is the estimated value.

Plot of predicted liquid heat capacity against experimental values is shown in Figure 1. The residuals of the MLR calculated values of the C_p are plotted against the experimental values in Figure 2. The propagation of the residuals on both sides of zero indicates that no systematic error exists in the development of the MLR.

Vibrating jumps are the most important factor for heat capacity determination in liquid compounds. Indeed, the important role of intermolecular forces must be considered. Frequency of vibration is in relation with the coefficient of force, the stronger the binding between atoms, the greater the coefficient of force will be. Consequently,

vibration locates at higher frequencies. In this case, while vibrating temperature increases, the heat capacity decreases.

Electronic descriptor of model, CNC, is a characteristic of intermolecular electronic interactions. Intermolecular forces became grater by increasing of partial charge over most negative atoms of carbon, and the heat capacity will increase. Four other descriptors show the influence of branches. While the more number of branches increase, force constant and consequently frequency of vibration becomes smaller. Hence, many number of molecules state in excited state and the heat capacity become greater. Good prediction ability of model, high correlation coefficient and root mean square error of 10.86 show that MLR is a convenient method for prediction of heat capa-
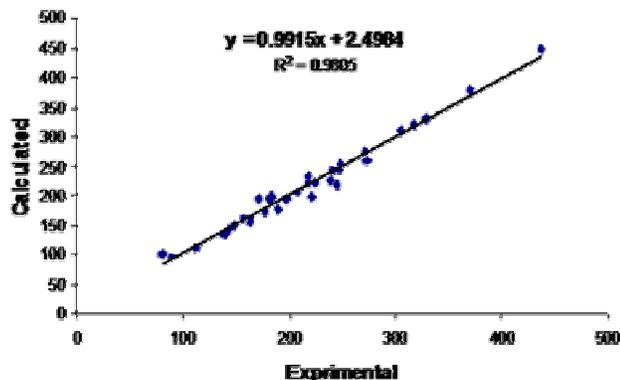
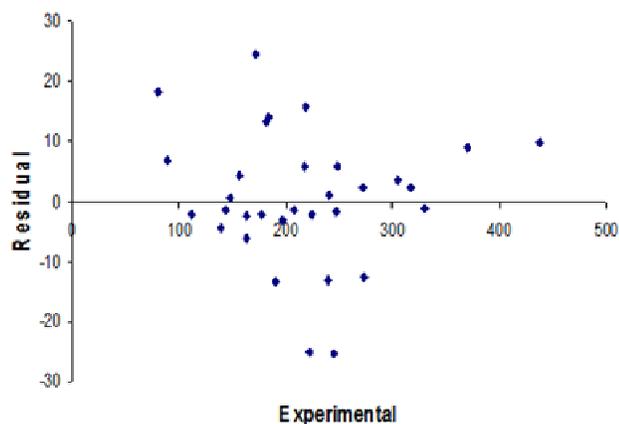**Figure 1.** Plot of calculated heat capacity against experimental heat capacity using MLR.

Figure 2. Plot of residuals against experimental value.

capacity of alcohols and aldehydes. The number of branches and polarity of molecules are more important than structural properties which affect on liquid heat capacity of alcohols and aldehydes.

REFERENCES

Balaban AT (1981). Chem. Phys. Lett. 89: 399.
Basac SC, Harriss DK, Magnuson VR (1984). J. Pharm. Sci. 73: 429.
David RL (1998) CRC handbook of chemistry and physics, 79th ED, 5.4 – 5.60.
Gakh AA, Gakh EG, Sumpter BG, Noid DW (1994). Neural network – Graph Theory approach to the prediction of the physical properties of organic compounds; J. Chem. Inf. Comput. Sci. 34: 832.
Ivanova AA, Palyulin VA, Zefirov AN, Zefirov NS, (2004). QSPR: Application to Heat Capacity Calculation; Russian J. Organic Chemistry, 40 (5): 644.
Kier LB (1980). J. Pharm. Sci. 69: 807.
Kier LB (1985). Quant. Struct. Act. Relat. 4: 109.
Kier LB, Hall LH (1981). J. Pharm. Sci. 70: 583.
Liu S, Cai S, Cao C, Li Z (2000). Molecular electronegative distance vector (MEDV) related to 15 properties of alkanes; J. Chem. Inf. Comput. Sci. 40: 1337.
Marija J (1992). Spss for windows, Ver. 5, Norusis / Spss Inc.
Moppac pakage (1995). ver. 4.0, US Air force acadey, Colorado spring Co. 80840.
Randic M (1975). J. Am. Chem. Soc. 97: 6609.
Rohrbaugh RH, Jurs PC (1987). Description of molecular shape applied in studies of structure / activity and structure / property relationships; J. Anal. Chem. Acta. 199: 99.
Roy K, Saha A (2006). QSPR with TAU indices. Part5, Liquid heat capacity of diverse functional organic compounds; J. Indian Chem. Society, 83: 351.
Stouch TR, Jurs PC (1986). Simple Method for the Representation, Quantification and Comparison of the Volumes and Shapes of Chemical Compounds; J. Chem. Inf. Comput. Sci. 26: 5.
Wiener H (1947). J. Am. Chem. Soc. 69: 17.
Yao XJ, Fan B, Doucet JP, Panaye A, Liu M, Zhang R, Zhang X, Hu Z (2003). Quantitative structure property relationship models for the prediction of liquid heat capacity; QSAR Comb. Sci. 22: 29.