Reaktor, Vol. 12 No. 4, Desember 2009, Hal. 260-267

A QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIP (QSPR) EVALUATION OF CRITICAL VOLUME OF UNSATURATED HYDROCARBON ALKENES AND ALKYNES USING SIMPLE CONNECTIVITY INDICES

Oman Zuas[*)] and Dyah Styarini

Research Centre for Chemistry – Indonesian Institute of Sciences (Pusat Penelitian Kimia - LIPI)
Kawasan PUSPIPTEK Serpong 15314, Tangerang, Banten, Indonesia.
Tel: +62-21-7560929, Fax: +62-21-7560549,
[*)]Corresponding author: oman.zuas@lipi.go.id

Abstract

A quantitative structure-property relationships (QSPR) was used in this study to relate the critical volume (Vc) of unsaturated hydrocarbons alkenes and alkynes compounds to their molecular structures. A QSPR study of Vc was performed on the basis of simple connectivity indices (SCI's). The obtained QSPR model is predictive and requires only one SCI descriptor in the calculation with statistical parameters including standard coefficient correlation (R^2)=0.997, cross-validated correlation coefficients (Q^2)=0.976, and average absolute error (AAE)=0.12. Application of the best QSPR model to a testing set of 30 alkenes and alkynes demonstrates good predictability without the needs in any experimental physicochemical properties data.

Key words*: alkenes, alkynes, connectivity indices, critical volume, QSPR, unsaturated hydrocarbons*

Abstrak

Pada kajian ini, hubungan kuantitatif sifat struktur (QSPR) suatu senyawa kimia digunakan untuk mengkorelasikan antara nilai volum kritis hidrokarbon tak jenuh kelompok alkena dan alkina dengan struktur molekulnya. Studi QSPR dilakukan berdasarkan diskriptor indeks konektifitas sederhana (SCI's) pada struktur molekul. Dari hasil perhitungan menunjukkan bahwa model QSPR yang dusulkan mempunyai kemampuan memprediksi nilai volum kritis senyawa alkena dan alkina dan membutuhkan hanya satu SCI dengan parameter statistik meliputi koefisien korelasi standar (R^2)=0,997, koefisian korelasi silang tervalidasi (Q^2)=0,976, and kesalahan absolut rata-rata (AAE)=0,12. Penerapan model QSPR terbaik terhadap 30 senyawa alkena dan alkina lainnya memperlihatkan bahwa model tersebut menghasilkan hitungan volume kritis yang baik tanpa membutuhkan data eksperimen.

Kata kunci: *alkena, alkina, indeks konektifitas, volum kritis, QSPR, hidrokarobon tak jenuh.*

INTRODUCTION

Critical volume (V*c*), one of critical properties of a chemical substance, is defined as the volume occupied by a certain mass, usually one gram molecule of a liquid or gaseous substance at its critical point (Filippov, 2004). This critical property is very important properties in chemical engineering fields because other thermo-chemical properties of substance such as critical temperature and critical pressure are predictable from critical volume by using van der Waals equation of state (Boynton and Bramley, 1992). Precise prediction of critical volume, therefore, is much needed.

Experimental measurements of some critical volume involve experimental difficulties and they are not always feasible, and the corresponding methods possess real drawbacks. Consequently, it is necessary to develop a theoretical calculation using computational-based technique, so called quantitative structure property relationship (QSPR), for calculating the critical volume of compound. (Karelson and Lobanov, 1996; Gute and Basak, 1997; Ribeiro and Ferreira, 2003; Erös *et al.,* 2004; Ghasemi and Saaidpour, 2007). A QSPR uses chemometric methods to describe how a given physical-chemistry properties of interest as a function of molecular descriptors describing the chemical structure of the molecule (Wold and Eriksson, 1995; Karelson and

Lobanov, 1996; Ribeiro and Ferreira, 2003). As a consequence, one of the most important points is the selection of adequate descriptors containing the information stored in the molecular structure. These descriptors are numerical representations of structural features of molecules that attempt to encode important information that causes structurally different compounds to have different physical property values (Wold and Eriksson, 1995; Gute and Basak, 1997).

Many methods estimating the critical volume of pure compound have been proposed in literature, and the group-contribution approach was generally used. In this technique, it is assumed that some property is a function of the molecule structure, *e.g*. the number and types of chosen molecular structures each of which is assigned a numerical value. Improvements of the methods have been well achieved by some authors by which allow describing the properties of various molecular structures (Constantinou and Gani, 1995; Marrero and Gani, 2001). In the same context, therefore, our study aims to evaluate structure properties on the basis of simple connectivity indices (SCI's) to estimate critical volume, V*c* (cm^3/mol) of pure unsaturated hydrocarbons focusing on a set of alkenes and alkynes compounds.

MATERIALS AND METHODS

Brief Description of the Molecular Descriptors Used

The connectivity indices have been widely used as molecular structural descriptors to correlate the physical properties of organic chemicals and use them in computational molecular design studies (Wang *et al*, 2006). Recently, the connectivity indices have been demonstrated the advantage of incorporating effects that are due to larger-scale structural features in a molecule on physical properties (Kier and Hall, 1986). In this work, we use six simple connectivity indices (Connectivity index Chi-0 through Chi-5) as listed in Table 1.

Table 1. The notation of the SCI's involved in the QSPR model

Notation	Descriptors
$^0\chi$	connectivity index chi-0
$^1\chi$	connectivity index chi-1 (Randic connectivity index)
$^2\chi$	connectivity index chi-2
$^3\chi$	connectivity index chi-3
$^4\chi$	connectivity index chi-4
$^5\chi$	connectivity index chi-5

The simple connectivity indices will further be called "SCI's" for short, is defined and developed by Kier and Hall (1986) and Randic (1975) that are calculated for the Hydrogen-depleted molecular graph. The SCI's provide quantitative characterization of skeletal variation in a molecule and are based on substructure features in the molecular graph, such as bonds, clusters and rings. Each feature is weighted according to number of skeletal neighbors for each atom, the connectivity simple delta value, δ. This simple δ value of an atom equals the number of neighboring atoms in the molecular skeleton and is subsequently used in calculating the SCI's (Randi´c, 1975; Kier and Hall, 1986). The SCI's used in this work which can generally be expressed by the following equation.

$$m_{\chi_t} = \sum_{i-1}^{N_S}\left[\prod_{k=1}^{m+1}\left(\delta_k\right)\right]^{-0.5} \quad (1)$$

Here m_{χ_t} is the molecular connectivity index of order *m* corresponding to sub graph type *t* (paths (P), clusters (C), or chain (CH)); N_S is the number of the relevant paths. In this work, for each chemical, the values of the connectivity indices up to fifth order are calculated from the vertex degree of the atoms in the Hydrogen-depleted molecular graph. The detailed equations for the SCI's for zeroeth, through fifth orders are defined as follows (Randi´c, 1975; Kier and Hall, 1986).

$$^0\chi = \sum_{i=1}\delta_i^{-0.5} \quad (2)$$

where, $^0\chi$ is connectivity index for zeroeth order, *n* is the number of nodes in the Hydrogen-depleted graph, δ_i is the vertex degree of the i^{th} atom defined as the number of non-Hydrogen neighbors in the molecular graph.

$$^1\chi = \sum_b\left(\delta_i\delta_j\right) \quad (3)$$

where, $^1\chi$ is connectivity index of the first order, *b* is the number of bonds, the sum runs through all bond in the Hydrogen-depleted molecule, and for each bond $\delta_i\delta_j$ is the product of the vertex degrees of the end atoms *i* and *j*.

$$m_\chi = \sum_{l=1}^{L}\left(\Pi\delta_i\right)_k^{-0.5} \quad (4)$$

where, m_χ is the SCI's of the m^{th} order for $2 \le m \le 5$, $\left(\Pi\delta_i\right)_k$ is the product of the vertex degrees of the atoms that form a connected sub graph with *m* edges, and L is the total number of such distinct connected sub graphs (the H-depleted molecular graph) each having *m* edges.

Data Set

A set of critical volume (V*c*) data of unsaturated aliphatic hydrocarbon (35 compounds) that has collected from Handbook of Chemical Compound Data for Process Safety by Yaws (Yaws, 1997) was adopted as training set. To test the predictive ability of the proposed model, the V*c* data for 30 unsaturated hydrocarbons collected from the literature (Yaws, 1997; Plyasunova, *et al.,* 2004) were investigated and adopted as the testing set. Both the training and the testing set consist of alkenes and

alkynes compounds as shown in Table 2 and were then used as dependent variables in the QSPR study.

Descriptors Calculations

All molecules were constructed using ChemDraw Ultra version 5.0 (CambridgeSoft, 1999) and it was saved as the MolFile (MDL MolFile) file format. For every compound, the MolFile (MDL MolFile) file format was opened with HyperChem Release 7.0 for windows (Hypercube, 2002), saved to HyperChem (HIN) file format to create the 3-Dimension model. Finally the model was subjected to calculation the SCI's descriptor by implementing the HIN files format on the DRAGON version 5.4 software (Telete, 2006). The SCI's descriptors were used as independent variables in the modeling.

Variables Selection

The data matrix X(y x z) with y = 65 rows and z = 6 columns corresponds, respectively, to the number of molecules investigated and molecular descriptors calculated (Table 3). This diversity of molecular descriptors was evaluated in order to find those that provide the best regression model for V*c*.

Modeling and Predictions

After the calculation of SCI's descriptors (Table 3), the QSPR model was build by means of the SPSS Release 13.0 for Windows using stepwise method. The highest correlation of independent variables with dependent variable was chosen for deriving the QSPR model. The classical QSPR regression equation can be obtained by the use of the scaled regression coefficients, mean and standard deviation of each original descriptor. The statistical parameters used to assess the quality of the models are the standard correlation coefficient (R^2) (Eq. 5), cross validated correlation coefficients (Q^2) (Eqs. 6), and the average absolute error (AAE) (Eqs. 7). The best model derived from the multiple linear regression (MLR) analysis was used to predict the V*c* of the testing set compounds (Table 2) which were not included in the training set.

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{n} \left(y_i - \overset{\Lambda}{y} \right)^2 \text{experimental}}{\sum_{i=1}^{n} \left[\left(y_i - \overline{y} \right)^2 \right]} \right] \qquad (5)$$

$$Q^2 = 1 - \left[\frac{\sum_{i=1}^{n} \left(y_i - \overset{\Lambda}{y} \right)^2}{\sum_{i=1}^{n} \left[\left(y_i - \overline{y} \right)^2 \right]} \right] \qquad (6)$$

In these equations, *n* is the number of compounds used for cross-validation, y^i is the experimental value of the physicochemical property for the i^{th} sample. $\hat{y}_i$ is the value predicted by the model built without sample *i*. $\overline{Y}$ is the mean value of experimental physicochemical property.

The average absolute error (AAE) (Eq. 7) was calculated as the following equation.

$$AAE = \frac{\Sigma \left| \left(Vc_{predicted} \right) - \left(Vc_{experimental} \right) \right|}{n} \qquad (7)$$

Where $Vc_{predicted}$ are predicted values of the V*c*, $Vc_{experimental}$ are the experimental values of the V*c*, and n is number of compounds.

RESULTS AND DISCUSSIONS

The V*c* experimental data of 35 unsaturated hydrocarbons consisting of alkenes and alkynes compounds listed as training set in Table 2 were used to construct the regression models and set as dependent variable. Six SCI's descriptors (Table 3) were set as independent variables. All possible combinations of the SCI's descriptors were investigated in order to obtain the best QSPR model. The stepwise MLR was used to select each independent variable for deriving a QSPR model by considering the correlation between each variable with the dependent variable. The MLR equation used for the QSPR model developed is as follows:

$$Y = a_1 b_1 + a_2 b_2 + a_3 b_3 \ldots\ldots\ldots a_n b_n + c \qquad (8)$$

Where Y is dependent variable. $a_1+a_2+a_3\ldots..a_n$ are the regression coefficients of independent variables. $b_1+b_2+b_3\ldots..b_n$ are independent variables. *C* is the regression constant obtained from the model fit.

Of six descriptors (Table 3), one descriptor have been automatically selected to model V*c* is ${}^0\chi$. The ${}^0\chi$ reflects the size of the molecule is the most significant descriptor that reflects the contribution of clusters in a molecule to critical volume. It should be mentioned that more models were obtained from the MLR analysis, but they were ruled out by the stepwise MLR procedure. The general purpose of MLR is to quantify the relationship between several independent variables (SCI's descriptors) and a dependent variable (V_c) of alkenes and alkynes compounds. A set of coefficients defines the single linear combination of independent variables that best describes critical volume of alkenes and alkynes compounds. To avoid self correlation between the variables used for the derivation of the QSPR model, the correlation matrix was calculated and the result shown in Table 4. The best QSPR model obtained from the MLR analysis is shown in Equation 9.

$$Vc = 79.198(\pm 0.789)^0\chi - 41.723(\pm 4.723) \qquad (9)$$

(n = 35; R^2 = 0.997; Q^2 = 0.976; *F* = 10,075; s = 11.104)

Table 2. Critical volume (V*c*) data for 65 alkenes and alkynes compounds involved in this work.

Training set					
Cpd. No	CAS. No	Cpd. Name	Critical volume (cm^3/mol)		Validation Error (%)
			Experimental	Calculated	
1	74-86-2	acetylene	113.00	116.62	-3.10
2	74-99-7	methyl acetylene	164.00	168.85	-2.87
3	115-07-1	propylene	181.00	172.85	4.72
4	590-19-2	1,2-butadiene	219.00	228.29	-4.07
5	107-00-6	ethyl acetylene	222.00	228.29	-2.75
6	106-99-0	1,3-butadiene	220.80	228.29	-3.28
7	106-98-9	1-butene	239.90	234.29	2.39
8	115-11-7	isobutene	238.90	241.75	-1.18
9	563-46-2	2-methyl-1-butene	292.00	297.19	-1.75
10	109-67-1	1-pentene	296.00	284.52	4.04
11	563-45-1	3-methyl-1-butene	302.10	297.19	1.65
12	598-25-4	3-methyl-1,2-butadiene	291.00	297.19	-2.08
13	591-95-7	1,2-pentadiene	276.00	284.52	-2.99
14	2004-70-8	trans-1,3-pentadiene	276.00	284.52	-2.99
15	591-96-8	2,3-pentadiene	295.00	284.52	3.68
16	627-19-0	1-pentyne	277.00	284.52	-2.64
17	760-21-4	2-ethyl-1-butene	364.00	353.42	2.99
18	763-29-1	2-Methyl-1-penten	359.00	353.42	1.58
19	760-20-3	3-methyl-1-penten	343.30	353.42	-2.86
20	00674-76-0	4-methyl-trans-2-pentene	346.00	353.42	-2.10
21	26519-91-5	Methyl cyclopentadiene	318.90	307.49	3.71
22	592-76-7	1-heptene	413.00	396.98	4.04
23	6094-02-6	2-methyl-1-hexene	398.00	409.65	-2.84
24	4038-04-4	3-ethyl-1-pentene	398.00	409.65	-2.84
25	3769-23-1	4-methyl-1-hexene	398.00	409.65	-2.84
26	111-66-0	1-octene	472.00	452.42	4.33
27	124-11-8	1-nonene	528.00	508.65	3.80
28	764-93-2	1-decyne	557.50	564.88	-1.31
29	872-05-9	1-decene	585.00	564.88	3.56
30	2243-98-3	1-undecyne	613.50	620.32	-1.10
31	821-95-4	1-undecene	642.00	620.32	3.50
32	765-03-7	1-dodecyne	669.50	676.55	-1.04
33	26186-02-7	1-Tridecyne	725.00	732.78	-1.06
34	765-13-9	1-pentadecyne	837.00	844.45	-0.88
35	629-74-3	hexadecyne	893.00	900.68	-0.85
Testing set					
Cpd. No	CAS. No	Cpd. Name	Critical volume (cm^3/mol)		Validation Error (%)
			Experimental	Predicted	
36	74-99-7	Propyne	163.50	170.61	-4.17
37	107-00-6	1-Butyne	220.00	228.60	-3.91
38	503-17-3	2-Butyne	221.00	228.60	-3.44
39	9003-27-4	2-Methyl-1-propene	238.80	241.51	-1.14
40	590-18-1	cis-2-butene	234.00	228.60	2.31
41	109-68-2	2-Pentene	292.40	284.60	2.67
42	591-93-5	1,4-Pentadiene	276.00	284.60	-3.11
43	627-19-0	1-Pentyne	278.00	284.60	-2.37
44	592-42-7	1,5-Hexadiene	328.00	340.59	-3.84
45	110-83-8	Cyclohexene	296.88	294.26	0.88
46	592-41-6	1-Hexene	350.00	340.59	2.69
47	563-78-0	2,3-dimethyl-1-butene	349.00	366.88	-4.87
48	7688-21-3	cis-2-Hexene	351.00	340.59	2.97
49	4050-45-7	trans-2-Hexene	351.00	340.59	2.97
50	7642-09-3	cis-3-Hexene	350.00	340.59	2.69
51	13269-52-8	trans-3-Hexene	350.00	340.59	2.69
52	563-79-1	2,3-Dimethyl-2-butene	351.00	366.49	-4.41

53	691-37-2	4-Methyl-1-pentene	353.40	353.58	-0.05
54	3070-53-9	1,6-Heptadiene	390.00	396.66	-1.71
55	628-71-7	1-Heptyne	386.00	396.66	-2.76
56	1119-65-9	2-Heptyne	390.00	396.66	-1.71
57	591-49-1	1-Methylcyclohexene	348.00	363.16	-4.36
58	6443-92-1	2-Heptene	406.00	396.66	2.30
59	629-05-0	1-Octyne	441.00	452.66	-2.64
60	3452-09-3	1-Nonyne	513.00	508.65	0.85
61	112-41-4	1-dodecene	700.00	676.55	3.35
66	2437-56-1	1-Tridecene	756.00	732.78	3.07
63	1120-36-1	1-tetradecene	817.00	789.01	3.43
64	13360-61-7	1-pentadecene	875.00	844.45	3.49
65	629-73-2	1-hexadecene	933.00	900.68	3.46

Table 3. SCI's descriptors selected to construct the QSPR models

Cpd. No	Cpd Name	SCI's descriptors					
		$^{0}\chi$	$^{1}\chi$	$^{6}\chi$	$^{3}\chi$	$^{4}\chi$	$^{5}\chi$
1	acetylene	2.00	1.00	0.00	0.00	0.00	0.00
2	Methyl acetylene	2.71	1.41	0.71	0.00	0.00	0.00
3	propylene	2.71	1.41	0.71	0.00	0.00	0.00
4	1,2-butadiene	3.41	1.91	1.00	0.50	0.00	0.00
5	ethylacetylene	3.41	1.91	1.00	0.50	0.00	0.00
6	1,3-butadiene	3.41	1.91	1.00	0.50	0.00	0.00
7	1-butene	3.41	1.91	1.00	0.50	0.00	0.00
8	isobutene	3.58	1.73	1.73	0.00	0.00	0.00
9	2-methyl-1-butene	4.28	2.27	1.80	0.82	0.00	0.00
10	1-pentene	4.12	2.41	1.35	0.71	0.35	0.00
11	3-methyl-1-butene	4.28	2.27	1.80	0.82	0.00	0.00
12	3-methyl-1,2-butadiene	4.28	2.27	1.80	0.82	0.00	0.00
13	1,2-pentadiene	4.12	2.41	1.35	0.71	0.35	0.00
14	trans-1,3-pentadiene	4.12	2.41	1.35	0.71	0.35	0.00
15	2,3-pentadiene	4.12	2.41	1.35	0.71	0.35	0.00
16	1-pentyne	4.12	2.41	1.35	0.71	0.35	0.00
17	2-ethyl-1-butene	4.99	2.81	1.92	1.39	0.29	0.00
18	2-Methyl-1-pentenene	4.99	2.77	2.18	0.87	0.58	0.00
19	3-methyl-1-pentenene	4.99	2.81	1.92	1.39	0.29	0.00
20	4-methyl-trans-2-	4.99	2.77	2.18	0.87	0.58	0.00
21	Methylcyclopentadiene	4.41	2.89	2.39	1.64	1.13	0.29
22	1-heptene	5.54	3.41	2.06	1.21	0.68	0.35
23	2-methyl-1-hexene	5.70	3.27	2.54	1.14	0.61	0.41
24	3-ethyl-1-pentene	5.70	3.35	2.09	1.73	0.87	0.00
25	4-methyl-1-hexene	5.70	3.31	2.30	1.48	0.70	0.20
26	1-octene	6.24	3.91	2.41	1.46	0.85	0.48
27	1-nonene	6.95	4.41	2.77	1.71	1.03	0.60
28	1-decyne	7.66	4.91	3.12	1.96	1.21	0.73
29	1-decene	7.66	4.91	3.12	1.96	1.21	0.73
30	1-undecyne	8.36	5.41	3.48	2.21	1.38	0.85
31	1-undecene	8.36	5.41	3.48	2.21	1.38	0.85
32	1-dodecyne	9.07	5.91	3.83	2.46	1.56	0.98
33	1-Tridecyne	9.78	6.41	4.18	2.71	1.74	1.10
34	1-pentadecyne	11.19	7.41	4.89	3.21	2.09	1.35
35	hexadecyne	11.90	7.91	5.24	3.46	2.27	1.48
36	Ethyne	2.00	1.00	0.00	0.00	0.00	0.00
37	Propyne	2.71	1.41	0.71	0.00	0.00	0.00
38	1-Butyne	3.41	1.91	1.00	0.50	0.00	0.00
39	2-Butyne	3.41	1.91	1.00	0.50	0.00	0.00
40	cis-2-butene	3.41	1.91	1.00	0.50	0.00	0.00
41	2-Methyl-1-propene	3.58	1.73	1.73	0.00	0.00	0.00
42	2-Pentene	4.12	2.41	1.35	0.71	0.35	0.00
43	1,4-Pentadiene	4.12	2.41	1.35	0.71	0.35	0.00

44	1-Pentyne	4.12	2.41	1.35	0.71	0.35	0.00
45	Cyclopentene	4.54	2.5	1.77	1.25	0.88	0.00
46	1,5-Hexadiene	4.83	2.91	1.71	0.96	0.50	0.25
47	Cyclohexene	4.24	3.00	2.12	1.5	1.06	0.75
48	1-Hexene	4.83	2.91	1.71	0.96	0.5	0.25
49	2,3-dimethyl-1-butene	5.16	2.64	2.49	1.33	0.00	0.00
50	cis-2-Hexene	4.83	2.91	1.71	0.96	0.50	0.25
51	trans-2-Hexene	4.83	2.91	1.71	0.96	0.50	0.25
52	cis-3-Hexene	4.83	2.91	1.71	0.96	0.50	0.25
53	6,3-Dimethyl-6-butene	5.16	2.64	2.49	1.33	0.00	0.00
54	4-Methyl-1-pentene	4.92	2.77	2.18	0.87	0.58	0.00
55	1,6-Heptadiene	5.54	3.41	2.06	1.21	0.68	0.35
56	6-Heptyne	5.54	3.41	2.06	1.21	0.68	0.35
57	1-Methylcyclohexene	5.11	3.39	2.74	1.89	1.31	0.09
58	Cycloheptene	4.95	3.5	2.48	1.75	1.24	0.89
59	1-Octyne	6.24	3.91	2.41	1.46	0.85	0.50
60	1-Nonyne	6.95	4.41	2.77	1.71	1.03	0.60
61	1-dodecene	9.07	5.91	3.83	2.46	1.56	0.98
66	1-Tridecene	9.78	6.41	4.18	2.71	1.74	1.10
63	1-tetradecene	10.49	6.91	4.53	2.96	1.91	1.23
64	1-pentadecene	11.19	7.41	4.89	3.21	2.09	1.35
65	1-hexadecene	11.9	7.91	5.24	3.46	2.27	1.48

Table 4. The correlation matrix of all SCI's descriptors used in QSPR studies

	V_c	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi$	$^4\chi$	$^5\chi$
V_c	1	0.998	0.997	0.980	0.969	0.957	0.955
$^0\chi$		1	0.997	0.982	0.969	0.956	0.953
$^1\chi$			1	0.974	0.971	0.968	0.966
$^2\chi$				1	0.958	0.944	0.927
$^3\chi$					1	0.954	0.907
$^4\chi$						1	0.941
$^5\chi$							1

Based on this relationship, the V*c* of 35 alkenes and alkynes compounds were predicted and the accuracy of the predictions was then assessed by the residuals between the experimental and predicted values. The experimental and predicted V*c* of alkenes and alkynes compounds in the training set based on the QSPR equation above are shown in Table 2 and the plot of both values is given in Fig. 1 and Fig. 2. The results indicate a good linear regression of correlation between experimental and predicted V*c*. The positive value of the coefficient for the $^0\chi$ descriptor implies that high size of the atom correlates with decreased the critical volume of the alkenes and alkynes molecules.

In previous study, Wang, *et al.* (2006) used molecular connectivity indices to develop a QSPR model for determining the aqueous solubility a set of chlorinated hydrocarbons compounds. They achieved in correlating the three connectivity indices descriptors that reflect the contribution of clusters in a molecule to aqueous solubility that are important in describing the aqueous solubility of chlorinated hydrocarbons compounds. Another study reported by Delgado (2002) demonstrated that CODESSA has been successfully in applying to develop QSPR model and carried out a correlation analysis to find the best QSPR model using a heuristic method. Delgado attained in obtaining the two descriptors that have definite physical meaning corresponding to different intermolecular interactions.

To test the predictive ability of our model, the V_c data for 30 alkenes and alkynes compounds were collected from the literature (Yaws, 1997; Plyasunov *et al.*, (2004) as the testing set. The predictive results calculated with equation (9) are shown in Table 2, where the experimental values and the residual values are also listed, and the scatter plot is shown in Fig 2.

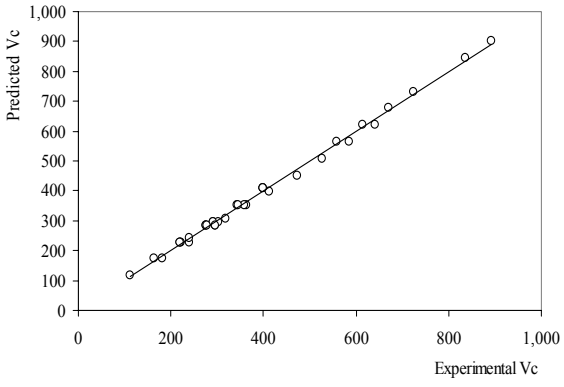

Figure 1. Plot of predicted V*c* values versus the experimental V*c* values of alkenes and alkynes in the training set.

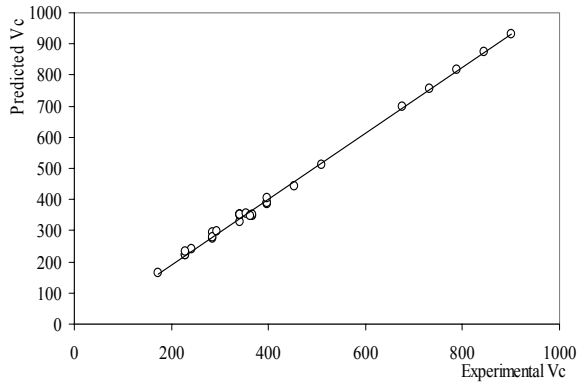

Figure 2. Plot of predicted V*c* values versus the experimental V*c* values of alkenes and alkynes in the testing set.

The prediction results are in good agreement with the experimental values. The low residual activity with high cross-validated coefficient correlation (Q^2 = 0.976) and low average absolute error (AAE = 0.12) observed indicates that the developed of QSPR model is reliable with good predictability.

CONCLUSIONS

Predictive QSPR model based on SCI's is proposed in this study to correlate the critical volume of unsaturated hydrocarbons compounds *i.e.,* alkenes and alkynes compounds. The application of the best model obtained to a testing set of 30 alkenes and alkynes compounds demonstrates that the new model is reliable with good predictability. The descriptors are able to describe the phenomena of the characterization of skeletal variation in a molecule and are based on substructure features in the molecular graph of the molecules. Besides, it was also possible to construct new model by applying SCI's approach without require any experimental physicochemical properties in the calculation of critical volume.

ACKNOWLEDGEMENTS

The author's thank go to Nandang, B.Sc. for assistance in gaining access to the Computational Centre at the Research Centre for Chemistry–LIPI, to Hery Kresnadi, M.Eng., for assistance in providing the valuable computational software.

REFERENCES

Boynton, W. P. and Bramley A., (1992), A modification of Van Der Waals' equation, *Phys. Review*. 20, pp. 46 - 50.

CambridgeSoft, (1999), CS Chem3D Pro version 5.0, Cambridge Soft Corporation. USA. *http://www.camsoft.com.* (accessed on March 7, 2007)

Constantinou, L., Gani, R., and O'Connel, J.P., (1995), Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method, *Fluid Phase Equilib.*, 103, pp. 11-22.

Delgado, E. J., (2002), Predicting aqueous solubility of chlorinated hydrocarbons from molecular structure, *Fluid. Phase. Equilibr.*, 199, pp. 101-107.

Erös, D., Kéria, G., Kövesdi, I., Szántai-Kis, C., Mészáros, G., and Örfi., (2004), Comparison of predictive ability of water solubility QSPR models generated by MLR, PLS and ANN methods: Mini-Review, *Med. Chem.*, 4, pp. 167-177.

Filippov, L. P., (2004), Calculation of the critical volume of substances based on the structural formula, *J. of Struct. Chem.*, 24, pp. 632-634.

Ghasemi, J. and Saaidpour, S., (2007), QSPR prediction of aqueous solubility of drug-like organic compounds, *Chem. Pharm. Bull.*, 55, pp. 669-674.

Gute, B. D. and Basak, S. C., (1997), Predicting acute toxicity of benzene derivatives using theoretical molecular descriptors: a hierarchical QSAR approach, *SAR and QSAR EnViron. Res.*, 7, pp. 117-131.

Hypercube, I., (2002), HyperChem Release 7.0 for Windows Molecular Modeling System, *www.hyper.com.* (Accessed on June 18, 2007)

Karelson, M. and Lobanov, V.S., (1996), Quantum-chemical descriptors in QSAR/QSPR studies, *Chemical reviews*, 96, pp. 1027-1043.

Kier, L. B. and Hall, L.H., (1986), *Molecular connectivity in structure-activity analysis,* 1st ed., RSP-Wiley, Chichetser. United Kingdom.

Marrero, J. and Gani, R., (2001), Group-contribution based estimation of pure component properties, *Fluid Phase Equilib.*, 184, pp. 183-208.

Plyasunova, N.V., Plyasunov, A.V., and Shock, E.L., (2004), Database of thermodynamic properties for aqueous organic compounds, *Intern. J. Thermophysic.* 25, pp. 351-360.

Randi´c, M., (1975), On characterization of molecular branching, *J. of American Chem. Soc.*, 97, pp. 6609-6615.

Ribeiro, F. A. D. and Ferreira, M. M. C., (2003), QSPR models of boiling point, octanol–water partition coefficient and retention time index of polycyclic aromatic hydrocarbons, *J. of Molec. Struct. (Theochem).*, 663, pp. 109-126.

Telete, S., (2006), DRAGON for Windows: software for molecular descriptor calculations version 5.4, *http://www.telete.mi.it.* (Accessed on February 6, 2008).

Wang, Y.L., Hu, L.D., Wu, L.Y., and An., W.Z., (2006), Predicting aqueous solubility of chlorinated hydrocarbons by the MCI approach, *International J. of Molec. Sci.*, 7, pp 47-58.

Wold, S. and Eriksson, L., (1995), *Chemometric Methods in Molecular Design*, 3rd ed., VCH, Weinheim. Germany

Yaws, C. L., (1997), *Handbook of Chemical Compound Data for Process Safety (Library of Physico-Chemical Property Data)*, 2nd ed., Gulf Professional Publishing, United Kingdom.