ISSN 1070-3632, Russian Journal of General Chemistry, 2014, Vol. 84, No. 11, pp. 2066–2072. © Pleiades Publishing, Ltd., 2014.
Original Russian Text © I.G. Zenkevich, A.N. Marinichev, 2014, published in Zhurnal Obshchei Khimii, 2014, Vol. 84, No. 11, pp. 1767–1773.

Estimation of Octane Rating of Hydrocarbon Homologs Using Logistic Regression

I. G. Zenkevich and A. N. Marinichev

St. Petersburg State University, Universitetskii pr. 26, St. Petersburg, 198504 Russia
e-mail: izenkevich@mail15.com

Received June 30, 2014

Abstract—Octane rating of hydrocarbons from engine fuels have been determined using the four-parametric logistic regression. Regression equations have been constructed for homologous groups, carbon atoms number being the independent variable. The method accuracy is no worse than that of experimental methods of octane rating determination.

Keywords: hydrocarbon, homolog, octane rating, approximation, logistic regression

DOI: 10.1134/S1070363214110024

Estimation of physico-chemical properties of related organic compounds (primarily in homologous groups) via regression methods is a topical issue [1] due to complications of experimental determination of some properties. Depending OR the raw data availability, the estimation methods are subdivided into the following groups:

(a) additive schemes utilizing the pre-determined increments of structural fragments of the studied molecules;

(b) methods utilizing the information OR change of the property in question in other, better studied, series of compounds;

(c) methods utilizing information OR change of other, better studied, properties in the series of compounds in question; and

(d) methods utilizing the known values of the property in question for the compounds belonging to the series under consideration.

The latter type of tasks presumes utilizing of regression methods in order to estimate the value A of the considered property accounting for the compound place in the considered series. The simplest case if the approximating equation of the $A(n)$ type with n being number of carbon atoms in the molecule.

The approximating functions $y = y(x)$ having two finite limits at $x \to 0$ and $x \to \infty$ form a special group of equations. Such equations result, for example, from processing kinetic data for auto-catalyzed reactions (the product acting as a catalyst) [2]. They can be fitted by four-parameter non-linear relation, so called logistic regression

$$y = \frac{a}{1 + be^{-kx}} + c \qquad (1)$$

having the following limits at $k \neq 0$:

$$\lim_{x \to 0} y = \frac{a}{1 + b} + c,$$

$$\lim_{x \to \infty} y = c,\ k < 0,$$

$$\lim_{x \to \infty} y = a + c,\ k > 0.$$

If parameters a and c are known, the $u = (y - c)/a$ (with $b \neq 0$) change of variables gives a single-parametric nonlinear differential equation (2), equivalent to Eq. (1).

$$\frac{du}{dx} = ku(1 - u). \qquad (2)$$

Logistic regression has widely applied in many tasks. For example, Eq. (1) describes kinetics of quercetin oxidation in aqueous solutions [3] and of solid-phase topochemical reaction of aryldiyne carbamates polymerization [4]; it has been applied to estimate the plants winter hardiness as well [5]. Processing of results of quantitative analysis (the sequential additives method) sometimes require extrapolation to the infinitely large additive, a possible solution of the problem is using logistic regression [6].

Examples of physico-chemical parameters of organic compounds that can be fitted with Eq. (1) within the homologous series have not been known so far. Herein we report OR determination of octane rating of hydrocarbons in several homologous groups taking advantage of the four-parametric logistic regression.

Octane rating (OR) reflects the detonation tolerance of engine fuel as well as of individual components of the fuel [7]. Octane rating can be experimentally determined using several methods [7, 8]; in particular, Motor Octane Rating MOR, Blending Octane Rating BOR, Pump Octane Rating POR, Road Octane Rating ROR, and Anti-Knock Index AKI are reported. Octane rating determined via different methods are generally close; however, exceptions are known: for example, in the case of 1,5-hexadiene MOR = 37.6 and BOR = 71.1. The difference between octane rating resulting from the determination method reflects sensitivity of the fuel to the changed regime of engine operation [8].

In order to estimate and compare the operational parameters of engine fuels and their components, the octane rating value should be reproducible using a standard method. At the same time, analysis of octane rating as a physico-chemical property of organic compounds is possible using mean values and standard deviations of the parameter determined using different methods. Result of statistical analysis of the reference values of octane rating for 34 hydrocarbons is given in Table 1.

Standard deviation of octane rating is generally below 10% of the mean value, being somewhat higher in the case of cycloalkanes. Hereafter we will discuss mean octane rating.

The experimental determination of octane rating is subject to certain complications; therefore, approaches to estimate OR from other physico-chemical properties have been developed. In particular, octane rating depends OR the ratio of numbers of hydrogen atoms at primary, secondary, and tertiary carbon atoms in a molecule; hence, OR can be estimated from IR spectroscopy [9–11] or Raman and ^{13}C NMR spectroscopy [12, 13] data. In the case of multicomponent fuels (but not for individual hydrocarbons) octane rating can be determined from data of gas chromatography analysis [14–17]. A method of optimization of fuel composition to obtain certain OR value has been proposed [18]. Other known methods to calculate octane rating of hydrocarbons utilize molecular topological connectivity indices [14], genetic algorithms [19], modified additive schemes [20, 21], and neural networks [13]. The trends of OR change within various homologous series (*n*-alkanes, 2-methylalkanes, and 2,2-dimethylalkanes) as well as in the series of congeners with varied number of methyl groups have been discussed in [22], but logistic regression has not been applied to the analysis of octane rating values.

The available reference data mainly include octane rating of petrol fraction hydrocarbons (those with normal boiling point below 220°C and containing up to 12 carbon atoms in the molecule). One-side estimations of OR are known for some compounds with large number of carbon atoms (for instance, OR < –30 for *n*-hexadecane), such estimations are not suitable for calculations. In general, simple compounds containing small number of carbon atoms have been well studied experimentally. However, octane rating have been experimentally determined for relatively few of numerous higher homologs. Hence, the problem of theoretical estimation of octane rating is of special importance as far as higher hydrocarbon homologs are considered. Further limitation of the choice of the compounds to be evaluated is that approximation with Eq. (1) requires at least four compounds of the group to be characterized experimentally. Taking the above-mentioned considerations into account, we selected 57 compounds (in 10 groups) of the available 264 compounds for analysis (Table 2). Note that (*E*)- and (*Z*)-isomers were included in the group of alkenes-2.

Successful approximation of octane rating values within groups of homologs requires that the approximating functions has a finite limit with increasing number of carbon atoms in the molecule; the validity of this assumption should be specially considered. Let us analyze a representative example of octane rating of *n*-alkanes C_1–C_{12} as function of the carbon atoms number in the molecule (Fig. 1).

Rigorous proof of existing of a finite limit at $n \to \infty$ may utilize a function with known conditions of finite limit to exist, well approximating the OR numbers. Linear recurrent equation (3) is a convent choice to describe monotonously changed physico-chemical properties A of organic compounds as function of the carbon atoms number n_C [23–25]; parameters p and q of Eq. (3) can be calculated using least squares method.

$$A(n_C + 1) = pA(n_C) + q. \quad (3)$$

Evidently, if in Eq. (3) $0 < p < 1$ then property A has a finite limit.

Table 1. Statistical processing of reference data OR octane rating of selected hydrocarbons

Compound	Experimental data OR octane rating	Mean octane rating ± standard deviation
Methane	120, 120, 110, 107.5	114±7
Ethane	108	108
Propane	112, 97, 100, 105.7, 97.1	102±6
n-Butane	90.1, 93.6, 91.0, 93, 113, 114, 94, 89.6, 93.8, 93.6	97±9
2-Methylpropane	102, 99.0, 101.1, 91, 122, 97.6, 110.1	103±10
Pentene-1	34, 77.1, 90.9, 90.0, 91, 77, 89, 76	84±7
Cyclopentane	101, 85, 141, 141, 100, 84.9, 98, 79, 101.3, 100.1	107±24
n-Pentane	62, 61.7, 61.9, 62, 62, 62, 67, 62.6, 61.7	62.6±1.8
2-Methylbutane	90.3, 92.3, 92, 90, 99, 104, 96, 93, 100	95±5
2,2-Dimethylpropane	85, 80, 100, 90, 85.5, 80.2	87±8
Benzene	101, 116.3, 113.0, 114.8, 106, >100, >100, 99, 91, 101, 93, 102.7, 105, 111.6, 102.8	104±8
Cyclohexane	77.2, 83.0, 83, 77, 110, 97, 82.5	87±12
Methylcyclopentane	80.0, 91.3, 91, 80, 107, 99, 74	89±12
n-Hexane	25, 26, 26, 24.8, 25, 26, 19, 22, 29, 33	26±4
2-Methylpentane	83, 73.4, 73.5, 75, 77, 74.5, 82	77±4
3-Methylpentane	86, 74.5, 74.3, 80, 78	79±5
2,2-Dimethylbutane	93.4, 91.8, 92, 93, 89, 97, 96, 93.4	93.2±2.5
2,3-Dimethylbutane	94.4, 96, 92.1, 105.8, 94.3, 103.5, 100.3	98±6
Toluene	121, 107, 114, 102.1, 115.7, 120.1, 103.5, 118, >100, >100, 124, 112, 114, 103, 111, 101, 100.3	112±8
Methylcyclohexane	75, 71, 104, 84, 70, 70, 74.8, 71.1	77±11
n-Heptane	0 (definition of octane rating), 4	0
2-Methylhexane	44, 44.6, 42.4, 46.6, 41, 46.4, 40	43.6±2.6
3-Methylhexane	55, 56, 52, 57, 59, 55.8	55.8±2.3
Ethylbenzene	107.4, 97.9, >100, 98, 124, 107, 112, 99, 97.9, 100.8	105±9
n-Octane	–17, –10, –19, –15, –20, –19, –19	–17±4
2-Methylheptane	23, 23.8, 21.7, 13, 19, 29, 20.6	22±5
3-Methylheptane	26.8, 35, 30, 27	30±4
2,2,3-Trimethylpentane	109.6, 99.9, 105, 101.0, 100, 100, 105, 112, 106, 101.2	105±4
2,3,3-Trimethylpentane	101.6, 99.4, 103, 100, 106.1, 111, 102, 100.6	103±4
2,2,4-Trimethylpentane (isooctane)	100 (definition of octane rating), 96, 97	100
Isopropylbenzene	>100, 99, 132, 124, 112, 102, 99.3, 102.1	110±13
2.2,3,3-Tetramethylpentane	123, 116.8, 117, 92, 95.0, 103.6	108±13
n-Undecane	–35, –35	–35
n-Dodecane	–40, –40	–40

Table 2. Parameters of logistic regression (1) equation for hydrocarbons of different homologous groups

Group of homologs	Number of compounds (N)	Parameters of Eq. (1)				
		a	$b \times 10^3$	$-k$	$c \pm s(c)$	$\lvert\bar{\Delta}\rvert$ [b]
n-Alkanes	12	151.5	5.16	0.920	–36.0±2.5	2.6
2-Methylalkanes	7	95.4	0.14	1.353	10.3±3.0	1.6
2,2-Dimethylalkanes	7	45.9	0.00	2.296	45.6±3.1	2.2
2,3-Dimethylalkanes	4	40.0	0.02	1.475	62.7[a]	–[a]
2,(n – 1)-Dimethylalkanes	4	94.2	0.09	1.190	13.4[a]	–[a]
n-Alkylcyclopentanes	5	154.0	1.49	0.862	–32.0±14.6	0.7
n-Alkylcyclohexanes	4	79.6	0.00	2.056	9.1[a]	–[a]
Cycloalkanes	4	87.2	7.05	0.953	59.4[a]	–[a]
Alkenes-1	6	146.2	4.14	0.659	–48.8[c]	0.7
Alkenes-2	4	39.8	0.00	2.038	52.2[a]	–[a]

[a] If a group contains four compounds, error of approximation cannot be calculated, and the $c = y(\infty)$ value is given without standard deviation. [b] $\lvert\bar{\Delta}\rvert$ is averaged absolute deviation of the calculated and the accepted experimental values of octane rating. [c] Reference data of octane rating gives the values in between 89 and 32, and the limiting value was thus estimated with huge error (–48.8±86).

$$A_\infty = \lim_{n_C \to \infty} A(n_C) = \frac{q}{1-p}.$$

As seen from the example in Fig. 1, the OR(n) dependences are nonlinear within the homologous groups. Such curves reveal an inflection point ($d^2A/dn^2 = 0$), and the plots of recurrent functions (3) contain two linear branches (Fig. 2).

Three points in the right-up part of the plot in Fig. 2 showed the almost linear dependence (3) for the C_1–C_4 n-alkanes. That part of the plot could be used to estimate the $\lim_{n\to 0}$OR(n) value, but octane rating for simplest alkanes are usually known from the experiment. The left-down linear part of the plot (for the C_5–C_{12} n-alkanes) is more interesting, reflecting extrapolation of the OR values to $n \to \infty$. The least-squares analysis of the data gave the following parameters of the linear plot [and, consequently, of Eq. (3)]: $p = 0.65 \pm 0.03$ and $q = -15.7 \pm 0.9$ ($R^2 = 0.992$) as $p < 1$, the finite limit of OR(n) at $n \to \infty$ exist, being equal to $-15.7/(1-0.65) \approx -44.9$.

Similar conclusion follows from analysis of more complicated second-order inhomogeneous difference equation (4) [26].

$$A(n_C + 2) = pA(n_C + 1) + qA(n_C) + r. \quad (4)$$

In contrast to Eq. (3), the Eq. (4) allows approximation of data for all the C_1–C_{12} alkanes with a single function. Solution of Eq. (4) can be written as Eq. (5)

$$A(n_C) - C_0 + C_1 z_1^n + C_2 z_2^n, \quad (5)$$

where z_1 and z_2 being the roots of the $z^2 = pz + q$ equation with parameters p and q following from analysis of Eq. (4) via least-squares method; parameters C_0, C_1, and C_2 result from the least-squares processing of Eq. (5) using z_1^n and z_2^n as basis functions [26].

Analysis of OR of the C_1–C_{12} n-alkanes using Eqs. (4) and (5) confirmed existence of the finite limit $\lim_{n\to 0}$OR(n) = –47.8, its value was close to that obtained using Eq. (3).

Hence, existence of a finite limit of n-alkanes OR for higher homologs was rigorously confirmed. Apparently, the same should be true for other homologous groups as well. Let us further consider the results of regression analysis of the data for various hydrocarbons.

Table 2 lists coefficients of Eq. (1) for ten homologous groups of hydrocarbons. Accuracy of the OR values approximation with the regression equation was characterized with average absolute deviation of

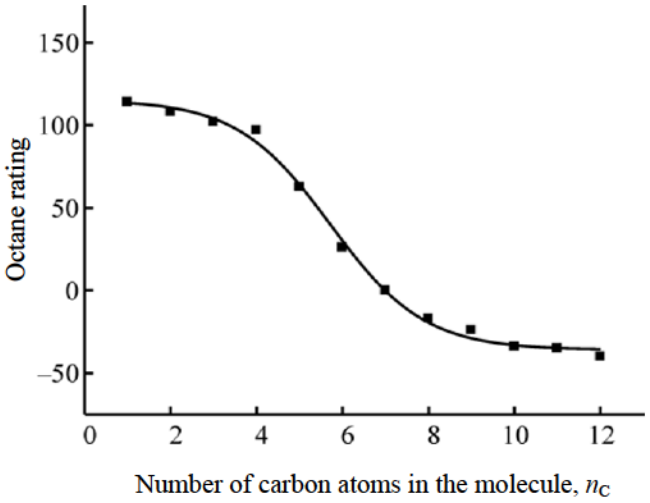

Fig. 1. Mean octane rating of *n*-alkanes C_1–C_{12} as func-tion of number of carbon atoms in the molecule. (Solid line) gives approximation with the logistic regression. Limiting values $y(0) = 115\pm3$ and $y(\infty) = -36\pm3$.

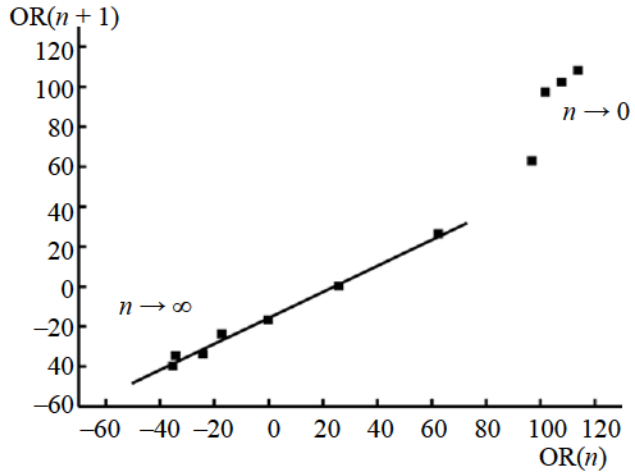

Fig. 2. Graphical presentation of recurrent function (3) for *n*-alkanes C_1–C_{12}.

the calculated OR from the corresponding mean experimental value (|Δ|). Five of the homologs groups contained only four compounds each, and accuracy of the approximation and the $c = \lim_{n\to0}OR(n)$ extrapolation could not be estimated. Nevertheless, even using the limited number of the hydrocarbons with reliably known OR we could convincingly demonstrate the potential of logistic regression analysis to approximate OR values for different classes of compounds. For example, solid line in Fig. 1 shows the logistic regression (1) approximation of octane rating values for 12 *n*-alkanes. The theoretical curve adequately describes the experimental values, the approximation error |Δ| being of 2.6. The limits of OR were $\lim_{n\to0}OR(n) = -115\pm3$ and $\lim_{n\to\infty}OR(n) = -36\pm3$. The first value was close to OR of methane (114), and the second estimation was in good agreement with the experimental value for *n*-dodecane as well as with recurrent estimations using Eqs. (3) and (4) (–45 and –48, respectively).

Let us consider estimations of the limiting OR values for hydrocarbons of different classes as obtained using the logistic regression. The $\lim_{n\to0}OR(n)$ values were in all the cases close, the average value being of 107±18 (104±11 after exclusion of the least reliable highest and lowest values). The reliable estimates of $\lim_{n\to\infty}OR(n)$ (with more than 4 compounds in the homologous group) revealed negative values for *n*-alkanes (–36), alkenes-1 (–49), and *n*-alkylcyclopentanes (–32). For the branched molecules the limiting values were higher: 10 (2-methyl-alkanes) and 46 (2,2-dimethylalkanes). The rough estimates (as calculated basing OR data for four homologs only) were as follows: 13 for 2,(n – 1)-dimethyl-alkanes (in agreement with the estimation for 2-methylalkanes, 10), –41 for 2-methylalkenes-1 (in agreement with the estimation for alkenes-1, –49), and 106 for the C_6–C_{10} *n*-alkylbenzene (the highest limiting OR throughout the studied classes).

To conclude, the $\lim_{n\to\infty}OR(n)$ limiting values reflected structural features of the studied homologs. The consequence of that is a limitation of logistic regression applications to estimate OR values: such regression analysis is not applicable for homologous groups containing less than four compounds, as using OR values for other classes of compounds to build the model may lead to significant inaccuracy of the estimation. Examples of such poorly characterized groups are dienes [with reference data available for 3-methyl-1,2-butadiene (52±13), 2-methyl-1,3-butadiene (90±13), and 1,5-hexadiene (scattered data)], cycloalkenes [cyclopentene (82±17) and cyclohexene (73±15)], and methyl *tert*-alkyl ethers [methyl *tert*-butyl ether 110±8) and methyl *tert*-amyl ether (106±12)]. Estimation of octane rating within these important classes required further development of the regression model.

Predictive performance of the logistic regression is additionally illustrated with data in Table 3, where the experimental values and these estimated using Eq. (1)

Table 3. Comparison of experimental and calculated octane rating of selected alkenes-1 and 2,(n – 1)-dimethylalkanes, along with the calculated values for the experimentally not characterized homologs

Number of carbon atoms in the molecule	Alkenes-1		2,(n – 1)-Dimethylalkanes	
	experiment[a]	calculated	experiment[a]	calculated
5	84	83	–	–
6	70	71	98	98
7	55	54	82	82
8	32	32	55.7	56
9	–	8	–	32
10	–	–12	20	20
11	–	–27	–	16
12	–	–36	–	14

[a] Mean of the experimental values determined using different methods.

are compared for homologs of alkenes-1 and 2,(n – 1)-dimethylalkanes, along with the estima-tions made for the compounds not studied experi-mentally. In all the cases deviations of the estimates from the corresponding experimental values were below unity.

EXPERIMENTAL

OR values used for the models construction were obtained by averaging the reference data from [16, 17, 22, 27, 28]. All the available types of OR (MOR, BOR, POR, ROR, and AKI) were averaged. Single values of OR were used without standard deviation, and the one-side estimations were excluded from calculations. Data for 249 hydrocarbons and of 15 other compounds were processed. The starting datasets and the averaged values of OR for selected compounds are given in Table 1. The logistic regression was built and analyzed taking advantage of Microsoft Excel 2007 ("Find Solution"), Microcal Origin 4.1/8.1 (Fit Sigmoidal), and Maple 14 (calculation of regression coefficients and estimation of the approximation accuracy) software.

REFERENCES

1. Karapet'yants, M.H., *Metody sravnitel'nogo rascheta fiziko-khimicheskikh svoistv* (Methods of Comparative Calculation of Physico-Chemical Properties), Moscow: Khimiya, 1965.
2. Denisov, E.T., *Kinetika gomogennykh reaktsii* (Kinetics of Homogeneous Reactions), Moscow: Vysshaya Shkola, 1988.
3. Gushhina, S.V., Kosman, V.M., and Zenkevich, I.G., *Vestn. S.-Peterburg. Gos. Univ., Ser. Fiz.-Khim.*, 2009, no. 1, p. 96.
4. Viazmin, S.Yu., Marinichev, A.N., and Domnin, I.N., *Russ. J. Appl. Chem.*, 2005, vol. 78, no. 10, p. 1662. DOI: 1070-4272/05/7810-1662.
5. Vasil'ev, N.P. and Egorov, A.A., *Matem. Biol. Bioinform.*, 2011, vol. 6, no. 2, p. 190.
6. Marinichev, A.N., Morozova, T.E., and Zenkevich, I.G., *Usp. Sovr. Estestvozn.*, 2013, no. 11, p. 152.
7. Gureev, A.A., Seregin, E.P., and Azev, V.S., *Kvalifikatsionnye metody ispytaniya neftyanykh topliv: Spravochnik* (Qualification Test Methods of Petroleum Fuels: Handbook), Moscow: Khimiya, 1984.
8. *Khimicheskaya entsiklopediya* (Chemical Encyclo-pedia), Knunyants, I.L., Ed., Moscow: Bol'shaya Sovetskaya Entsiklopediya, 1992, vol. 3, p. 367.
9. *Chemometrics Application Note N 11–6*, Infometrix Inc.: Bothell, WA. 1993.
10. Cao, D., Tan, J., Chen, Hu, and Han, S., *Guang Pu.*, 1999, vol. 19, no. 3, p. 314.
11. Chung, H., Lee, H., and Jun, C.-H., *Bull. Korean Chem. Soc.,* 2001, vol. 22, no. 1, p. 37.
12. Meusinger, R. and Moros, R., *Chemometr. Intel. Lab. Syst.*, 1999, vol. 46, no. 1, p. 67. DOI: 0169-7439/99/$.
13. Meusinger, R. and Moros, R., *Fuel*, 2001, vol. 80, p. 613. DOI: 0016-2361/01/$.
14. Anderson, P.C., Sharkey, J.M., and Walsh, R.P., *J. Instr. Petrol.*, 1972, vol. 59, p. 83.
15. van Leeuwen, J.A., Jonker, R.J., and Gill, R., *Chemometr. Intel. Lab. Syst.*, 1994, vol. 25, no. 2, p. 325. DOI: 0169-7439/94/$07.00.

16. Nikolaou, N., Papadopoulos, C.E., Gadlias, I.A., and Pitarakis, K.G., *Fuel*, 2004, vol. 83, no. 4, p. 517. DOI: 10.1016/j.fuel., 2003.09.011.
17. Ghosh, P., Hickey, J.H., and Jaffe, S.B., *Ind. Eng. Chem. Res*., 2006, vol. 45, p. 337. DOI: 10.1021/ie050811h.
18. Murty, B.S.N. and Rao, R.N., *Fuel Proc. Technol*., 2004, vol. 85, p. 1595. DOI: 10.1016/j.fuproc.2003.08.004.
19. Hosoya, H., *Croatica Chem. Acta*., 2002, vol. 75, no. 2, p. 433.
20. Albahri, T.A., *Fuel Chem. Div. Preprints*, 2002, vol. 47, no. 2, p. 531.
21. Albahri, T.A., *Ind. Eng. Chem. Res*., 2003, vol. 42, p. 657. DOI: 10.1021/ie020306+.
22. Perdih, A. and Perdih, F., *Acta Chim. Slov*., 2006, vol. 53, p. 306.
23. Zenkevich, I.G., *Russ. J. Struct. Chem*., 2007, vol. 48, no. 6, p. 1006. DOI: 10.1007/s10947-007-0164-z.
24. Zenkevich, I.G., *J. Chemometr*., 2010, vol. 24, p. 158. DOI: 10.1002/cem.1297.
25. Zenkevich, I.G., *Chromatography*, 2012, no. 75, p. 767. DOI: 10.1007/s10337-012-2260-2.
26. Marinichev, A.N., *Russ. J. Appl. Chem*., 2007, vol. 80, no. 2, p. 209. DOI: 10.1134/s1070427207020085.
27. Do, P.T.M., Crossley, S., Santikunaporu, M., and Resasco, D., *Catalysis*, 2007, vol. 20, p. 33. DOI: 10.1039/b602366p.
28. *http://www.astm.org/Standards/D2598.htm* (June, 2014).