Russian Journal of Applied Chemistry, Vol. 78, No. 3, 2005, pp. 453–455. Translated from Zhurnal Prikladnoi Khimii, Vol. 78, No. 3, 2005, pp. 457–460.

ORGANIC SYNTHESIS AND INDUSTRIAL ORGANIC CHEMISTRY

The Use of IR Spectra for Estimating the Boiling Points and Molecular Weights of Monohydric Aliphatic Alcohols

V. V. Vazhev and M. K. Aldabergenov

Kostanai State Teacher's Training College, Kostanai, Kazakhstan

Received December 9, 2004

Abstract—The use of transformed IR spectra as descriptors of molecular structure in estimating the boiling points and molecular weights of alcohols was studied. No additional data on molecular structure are required for the estimation.

The boiling point T_b is a fundamental characteristic of organic compounds; therefore, development of methods of its theoretical estimation attracts much attention for a long time. Numerous additive methods for estimation of T_b were suggested, but the most widely used is the QSPR (Quantitative Structure–Property Relationships) method, whose current status can be judged from published reviews [1, 2]. The majority of studies on estimating T_b concerns hydrocarbons or heterogeneous systems containing hydrocarbons and other low-polarity molecules. Much less attention was given to simulation and estimation of T_b of highly polar substances. This is apparently caused by the fact that the traditionally used descriptors of molecular structure (mainly topological indices) insufficiently adequately reflect the structural features of molecules responsible for formation of hydrogen bonds in solutions. In particular, the QSPR studies of T_b of alcohols were made only in a few papers [3–9]. The model constructed in [3] for a small set of 37 monohydric alcohols was characterized by the correlation coefficient between experimental and evaluated boiling points $R = 0.978$ and standard deviation $S = 7.4988$. For the same subjects, Yang *et al*. in a later study [4] obtained $R = 0.9837$ and $S = 6.35$. For the set of 29 alcohols Galvez *et al*. obtained $R = 0.979$ and $S = 3.63$ [5]. Recently [6] it was found that, for a small set of 54 saturated alcohols, $R = 0.9945$ and $S = 4.26$. For 58 alcohols, Randic and Basak [7] constructed a series of models, the best of which had $R = 0.9943$ and $S = 3.891$. For 106 monohydric alcohols, Cao and coworkers [8] constructed a model with $R = 0.9899$ and $S = 4.5727$. The highest quality of estimation with $R = 0.9960$ and $S = 3.814$ was reached by Ren for 108 alcohols [9].

Here we studied the possibility of estimating T_b and molecular weight of monohydric alcohols using their IR spectra as descriptors of molecular structure. Previously the IR spectra were successfully used in estimating chromatographic retention indices of a set of substances of various classes, among them alcohols [10], and normal boiling points of alkanes [11]. The possibility of determining the molecular weights of compounds from the IR spectra was not considered previously.

EXPERIMENTAL

We used the IR transmission spectra in the gas phase and the boiling points of monohydric aliphatic alcohols with the number of carbon atoms from 2 to 12, available from the server of the National Institute of Standards (NIST, USA) [12]. The training sample contained 60 substances, and the control sample, 40. Before the calculations, the spectra were normalized by the formula

$$d_{ij} = h_{ij}\left(\sum_{k=1}^{k=n} h_{ik}\right)^{-1}, \qquad (1)$$

where d_{ij} are operating values of descriptors used in calculations; n, number of points in the IR spectrum; and h_{ij}, transmission in the spectrum of substance i in point j.

A part of spectra in the range 550–3798 cm^{-1} was used; in this range, the spectrum was interpolated linearly with a step of 4 cm^{-1}; the number of points in the spectrum $n = 813$. Elements d_{ij} form a matrix of descriptors D in which submatrices of training sample A and control sample C can be distinguished. In our case, the number of descriptors (813) exceeds the number of substances in the training sample (60) and the problem of choice of the least correlated descriptors, well-known in estimation, arises. A new

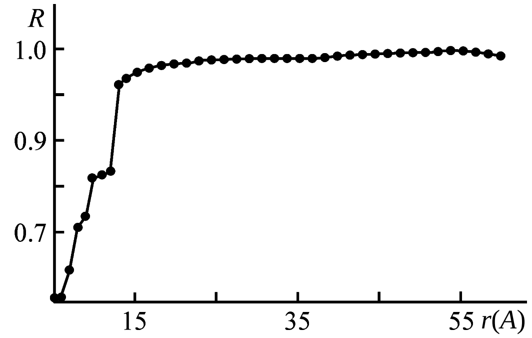

Fig. 1. Correlation coefficient R for the whole set of substances as a function of the estimated rank of descriptor matrix $r(A)$.

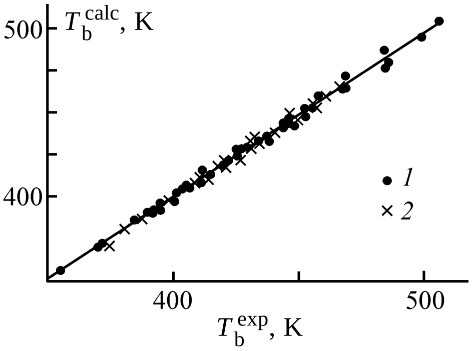

Fig. 2. Correlation between experimental boiling points of alcohols (T_b^{exp}) and those estimated from the IR spectra (T_b^{calc}). Sample: (*1*) training and (*2*) control; the same for Fig. 3.

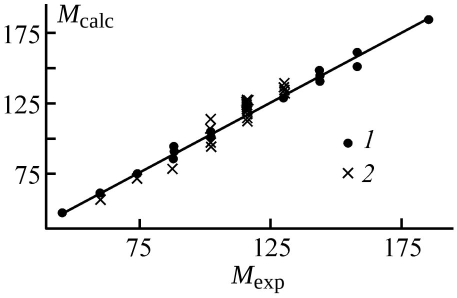

Fig. 3. Correlation between the experimental molecular weight M_{exp} and that estimated from the IR spectra (M_{calc}).

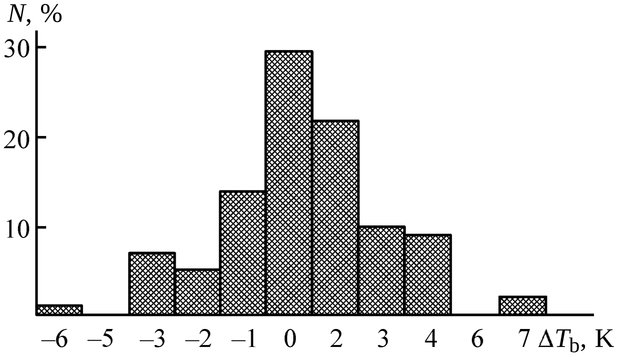

Fig. 4. Error distribution histogram for estimation of T_b of alcohols. (N) Incidence.

Table 1. Indexes of correlation between the experimental and calculated boiling points and molecular weights of alcohols

Characteristic	Index of correlation	Whole set	Sample	
			training	control
T_b	R	0.9972	0.9976	0.9963
	S	2.268	2.324	2.129
M	R	0.9896	0.9974	0.9693
	S	3.441	1.929	4.751

approach to solving of this problem was suggested in [13]. Here we present the expression used for the calculations:

$$p = CV_r\Gamma_r^{-1}U_r^{T}b, \qquad (2)$$

where p is the sought-for vector consisting of characteristics of the control sample; b, vector consisting of characteristics of the training sample of substances; U, matrix consisting of eigenvectors of matrix AA^T; V, matrix consisting of eigenvectors of matrix A^TA; and Γ, diagonal matrix of nonzero singular numbers coinciding for both matrix products.

The subscript is the estimated rank of the descriptor matrix $r(A)$; it denotes that r highest eigenvalues and the corresponding eigenvectors are used. For comparison of the results obtained with published data, the submatrix of descriptors in Eq. (2) was completed with the submatrix of training sample A and, correspondingly, vector b was completed with vector p. This procedure is equivalent to constructing a correlation model from the training sample and estimating the characteristics of both control and training samples. Calculations were performed using the PROGROC software (PROGram RObustness Calculation) developed by us [13]. The quality of estimation was characterized by the correlation coefficient R between the estimated and experimental values and by the standard deviation S.

The dependence of the correlation coefficient for T_b of the whole set of substances on the estimated rank of the matrix of descriptors is shown in Fig. 1.

The maximum $R = 0.9972$ for T_b is reached at $r = 54$, standard deviation $S = 2.268$. The results of estimation of T_b and molecular weight at this rank are shown in Figs. 2 and 3.

The quantitative characteristics of the corresponding correlation dependences are listed in Table 1.

A part of the results obtained in estimating T_b

Table 2. Experimental and estimated from the IR spectra boiling points of alcohols (K) for the control sample

Substance	T_b, K		Difference
	experiment	calculation	
2-Ethyl-1-pentanol	439.15	436.24	2.91
5-Methyl-3-hexanol	420.15	422.75	–2.60
5-Methyl-1-heptanol	455.65	454.30	1.35
2,3-Dimethyl-1-butanol	417.70	419.19	–1.49
2-Hexanol	411.00	412.53	–1.53
2,5-Dimethyl-2-hexanol	426.65	423.05	3.60
4-Methyl-2-ethyl-1-pentanol	449.65	445.96	3.69
2,2,3-Trimethyl-3-pentanol	426.65	423.38	3.27
4-Methyl-2-hexanol	421.15	420.35	0.80
2-Methy-1-propanol	380.80	381.25	–0.45
2,3-Dimethyl-3-pentanol	412.90	413.27	–0.37
2,3-Dimethyl-3-heptanol	446.15	450.46	–4.31
2-Ethyl-1-butanol	421.00	420.47	0.53
4-Methyl-3-hexanol	422.65	422.27	0.38
3-Pentanol	388.00	387.47	0.53
3-Hexanol	407.00	407.30	–0.30
6-Methyl-2-heptanol	445.00	447.16	–2.16
3-Heptanol	425.15	424.67	0.48
6-Methyl-1-heptanol	460.15	461.47	–1.32
3-Methyl-1-pentanol	426.00	425.32	0.68
2-Methyl-3-pentanol	398.00	397.93	0.01
4-Nonanol	465.65	466.18	–0.53
2-Methyl-2-butanol	375.10	370.98	4.12
5-Methyl-1-hexanol	440.65	439.31	1.34
2-Heptanol	432.00	436.49	–4.49
4-Methyl-4-heptanol	434.20	433.14	1.06
5-Methyl-2-hexanol	422.20	419.67	2.53
3-Methyl-3-hexanol	414.00	412.34	1.66
2-Methyl-2-hexanol	414.00	414.44	–0.41
Isopropanol	355.00	349.16	5.84
2-Methyl-2-heptanol	432.00	431.50	0.50
1-Hexanol	430.00	433.69	–3.69
2,2-Dimethyl-3-pentanol	409.00	409.50	–0.57
2-Methyl-3-ethyl-3-pentanol	432.15	432.70	–0.55
2-Ethyl-1-hexanol	456.00	456.11	–0.11
3-Methyl-2-pentanol	407.00	405.50	1.50
2,5-Dimethyl-3-hexanol	431.00	430.33	0.67
2-Octanol	453.00	454.14	–1.14
2,3-Dimethyl-2-pentanol	410.65	411.42	–0.77
4-Methyl-1-heptanol	455.85	454.04	1.81

of alcohols, related to the control sample only, are presented in Table 2.

The maximal error is observed for isopropanol and reaches 5.84 K. The error distribution for T_b of alcohols (Fig. 4) is close to the normal distribution, which confirms the reliability of the model; 71% of deviations of the calculated boiling points are in a narrow range (±2 K).

The above data show that our estimation of T_b of alcohols is better than the best published results obtained with the commonly used statistical characteristics.

CONCLUSIONS

(1) The use of normalized IR spectra as descriptors of molecular structure allows simultaneous estimation of the boiling point and molecular weight of alcohols in one numerical experiment.

(2) The quality of estimation is characterized by the correlation coefficient between the experimental and estimated boiling points $R = 0.9972$ and standard error of the estimate $S = 2.268$. For the molecular weight, $R = 0.9896$ and $S = 3.441$.

(3) The important feature of the suggested approach is the possibility of estimating the characteristics of unknown compounds without identification, using only their IR spectra.

REFERENCES

1. Katritzky, A.R., Maran, U., Lobanov, V.S., and Karelson, M., *J. Chem. Inf. Comput. Sci.*, 2000, vol. 40, no. 1, pp. 1–18.
2. Taskinen, J. and Yliruusi, J., *Adv. Drug Delivery Rev.*, 2003, vol. 55, pp. 1163–1183.
3. Yao, Y., Xu, L., Yang, Y., and Yuan, X., *J. Chem. Inf. Comput. Sci.*, 1993, vol. 33, no. 4, pp. 590–594.
4. Yang, Y.-O., Xu, L., and Hu, C.-Y., *J. Chem. Inf. Comput. Sci.*, 1994, vol. 34, no. 5, pp. 1140–1145.
5. Galvez, J., Garcia, R., Salabert, M.T., and Soler, R., *J. Chem. Inf. Comput. Sci.*, 1994, vol. 34, no. 3, pp. 520–525.
6. Kompany-Zareh, M., *Acta Chim. Slov.*, 2003, vol. 50, pp. 259–273.
7. Randic, M. and Basak, S.C., *J. Chem. Inf. Comput. Sci.*, 1999, vol. 39, no. 2, pp. 261–266.
8. Cao, C., Liu, S., and Li, Z., *J. Chem. Inf. Comput. Sci.*, 1999, vol. 39, no. 6, pp. 1105–1111.
9. Ren, B., *Comput. Chem.*, 2002, vol. 26, pp. 223–235.
10. Vazhev, V.V., *Vestn. Kaz. Nats. Univ. im. Al'Farabi* (Almaty), 2003, no. 3 (31), pp. 86–91.
11. Vazhev, V.V., *Neft' Gaz* (Almaty), 2003, no. 4, pp. 74–82.
12. *NIST Chemistry WebBook: NIST Standard Reference Data-base*, November 1998 Release, no. 69. http://webbook.nist.gov/chemistry/
13. Vazhev, V.V., *Ispol'zovanie IK- i mass-spektrov v QSAR/QSPR-issledovaniyakh* (The Use of IR and Mass Spectra in QSAR/QSPR Studies), Kostanai: Kost. Gos. Univ., 2003.