

PREDICTION OF BOILING POINTS OF ACYCLIC ALIPHATIC ALCOHOLS FROM THEIR STRUCTURE

Plamen Penchev, Nikolay Kochev, Vicktoria Vandeva, George Andreev
University of Plovdiv, Faculty of Chemistry,
Department of Analytical Chemistry

ABSTRACT

The normal boiling points for 120 C₅-C₈ aliphatic alcohols have been predicted using multiple linear regression analysis of different types of molecular descriptors derived from topology of the molecule and quantum chemistry calculations. Two new descriptors were introduced. The first one is the relative connectivity of oxygen atom expressed by the sum of all paths to the oxygen atom divided by Wiener index. The second one is a geometrical descriptor that reflects the oxygen atom shielding by spatially close atoms: it is a sum of the ratios of Van der Waals radius and distance from the oxygen to the corresponding atom raised to the third power.

Keywords: boiling point, multilinear regression, quantum chemistry, topology

INTRODUCTION

Boiling point (b.p.) is an important physicochemical property with practical value in chemistry, environmental protection and pharmaceutical industry. However, b.p. data often is not available, and therefore must be estimated theoretically. Estimation methods for b.p. have been widely explored [1-3] using topology of the molecule and/or quantum chemistry parameters calculated for optimized structure of the molecule.

Boiling point of a molecule depends on two major groups of factors. The first one includes intermolecular forces, such as dipole-dipole and Coulomb interactions. The second group accounts for the size and structure of the molecule as a whole, i.e. how the energy supplied by the heating is distributed into rotational and vibrational modes. That is why, every model for boiling point prediction has to account for these two trends with corresponding parameters.

In this work we present several approaches for theoretical calculation of normal boiling points of alcohols. The normal boiling points for 120 C₅-C₈ acyclic aliphatic alcohols have been predicted using multiple linear regression (MLR) analysis of different types of molecular descriptors: topological and quantum chemistry ones.

THEORETICAL MODELS

The topological parameters [4] used in different MLR models are given in Table 1. A new descriptor has been introduced designated as $W_{\text{Rel}}^{\text{O}}$. It is the relative connectivity of oxygen atom expressed by the sum of all paths to the oxygen atom divided by Wiener index.

Table 1. Topological parameters used in the MLR models

parameter	Parameter description
W	Wiener index – the sum of all path lengths in the molecule
$W_{\text{Rel}}^{\text{O}}$	The sum of all paths to the oxygen atoms divided by W
MW	Molecular weight
L_{C}	Size of the longest aliphatic chain
ξ	Molecular eccentricity index
χ^m	Connectivity index of order m (m = 1)
χ^{C}	Carbon connectivity - χ is calculated only for carbons
χ_{C}^m	Connectivity cluster index of order m (m = 3, 4)
χ_{P}^m	Connectivity path index of order m (m = 0, 1)
χ_{PC}^m	Connectivity path-cluster index of order m (m = 4, 5)
$\chi_{\text{PC}}^{\text{v}m}$	Valence connectivity path-cluster index of order m (m = 4, 5)
BCUT_{c}	The largest eigen value of the Burden matrix with weights based on partial charges
χ^m_{K}	Kier and Hall molecular shape indices of order m (m = 2)

The quantum chemistry calculations give the geometry of the molecule and some electronic parameters, such as dipole moment and partial charge of the atoms: the latter two determine the dipole-dipole and Coulomb interactions. For aliphatic acyclic alcohols it is the oxygen atom's charge that influences most the interaction. However, the strength of the interaction depends also on the oxygen surroundings which is quite different in various isomers of the alcohols. That is why, the new geometric parameter, so called oxygen shielding, O_{shield} , has been introduced.

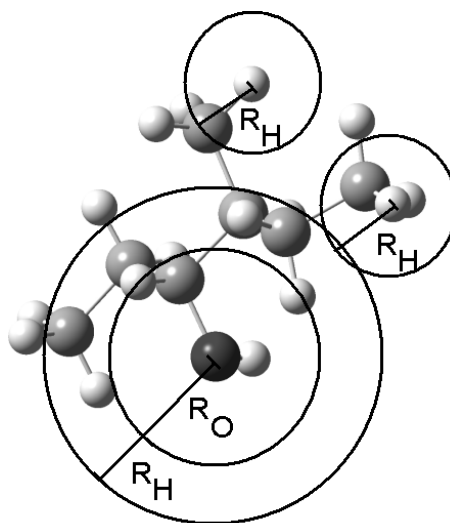


Figure 1. Illustration of oxygen shielding

O_{shield} is calculated by Equation (1) where the R_X is Van der Waals radius of the atom X and r_k is a distance of atom X_k ($X = \text{C}$ or H) to the oxygen: see figure 1 for clarity.

$$O_{\text{shield}} = \sum (R_{Xk} / r_k)^3; \quad (1)$$

the sum is taken only for atoms X_k for which $r_k < R_O + R_H + R_{Xk}$

The other quantum chemistry parameters tested in the models are the partial charge of the oxygen atom, q_O , partial charges of carbon, q_C , and hydrogen, q_H , that are connected to the oxygen and the dipole moment of the molecule, μ ; it has to be pointed out that the partial charge is not a measurable quantity and depends on scheme for its calculation.

The set of all parameters or a selection of them was used for MLR calculations and significance of the parameters was determined by stepwise model selection.

RESULTS AND DISCUSSION

The three quantum chemistry parameters, O_{shield} , q_O , and q_C , showed high correlation with boiling point when alcohols are separated in classes (see Table 2). This has to be expected, as these parameters account for intermolecular interactions but not how the energy supplied by the heating is distributed. It has to be mentioned that the value of q_C depends entirely on the type of the alcohol (primary, secondary, tertiary) and is little influenced by the other substituents in the molecule but this is not the case with q_O which depends on all substituents. As can be seen from the table, q_H badly correlates with b.p. and the dipole moment shows weaker correlation than the first three parameters and also does not correlate with b.p. of the whole set.

On the other side, ${}^1\chi$ parameter has a good descriptive power of the whole set – its correlation coefficient with b.p.s is 0.883. This topological parameter accounts for both the branching of molecule and its size: the first one strongly correlates with the oxygen shielding.

Table 2. The correlation coefficients between different parameters and b.p. calculated for different classes of isomers and for all alcohols

parameter	pentanols	hexanols	heptanols	octanols	all alcohols
O_{shield}	-0.906	-0.928	-0.909	-0.820	-0.325
q_O	0.891	0.864	0.849	0.682	0.296
q_C	-0.816	-0.910	-0.851	-0.759	-0.389
q_H	0.222	0.521	-0.041	0.102	0.078
μ	0.867	0.696	0.615	0.727	0.262
${}^1\chi$	0.848	0.794	0.758	0.717	0.883

Several different sets of descriptors were used in MRL model and a stepwise selection was performed upon them. The best seven descriptor sets are given in table 3.

Table 3. Best descriptor sets obtained by different MLR stepwise selections

Model #	Set of descriptors	Multiple R	Standard error (°C)
1	ln(W)	0.861	9.24
2	ln(W), $W_{\text{Rel}}^{\text{O}}$	0.958	5.18
3	$W_{\text{Rel}}^{\text{O}}$, $(^1\chi)^{1/2}$	0.982	3.39
4	$W_{\text{Rel}}^{\text{O}}$, $\ln(^1\chi)$, $^4\chi_{\text{pc}}^{\text{v}}$	0.986	2.98
5	$W_{\text{Rel}}^{\text{O}}$, $\ln(^1\chi)$, $^4\chi_{\text{pc}}^{\text{v}}$, O_{shield}	0.975	2.88
6	ln(W), $W_{\text{Rel}}^{\text{O}}$, $^0\chi_{\text{p}}$, $^1\chi_{\text{p}}$, $^5\chi_{\text{pc}}^{\text{v}}$	0.990	2.54
7	ln(W), $W_{\text{Rel}}^{\text{O}}$, MW, $^{\text{C}}\chi$, $^3\chi_{\text{c}}$, $^4\chi_{\text{c}}$, $^4\chi_{\text{pc}}$, $^4\chi_{\text{pc}}^{\text{v}}$	0.992	2.34

The two-, three and five-variable models (#3, #4 and #6) showed less standard error than the models described in Jurs paper [1] with the same number of variables. On the other side, their other models performed better than ours. It is interesting that O_{shield} appeared only in the four-variable model (#5) and it is the only quantum chemistry parameter that competed with the topological ones. It is the only parameter in this study that depends on the 3-D structure of the molecule, and an adverse consequence of this fact is its variability upon conformational changes in the molecule. We intend to study this adverse effect in future work.

The eight-variable model (#7) is given with the following equation:

$$\text{b.p.} = (43.71 \pm 10.05) \ln(W) + (134.04 \pm 17.96) W_{\text{Rel}}^{\text{O}} + (1.21 \pm 0.23) \text{MW} - (21.106.34) ^{\text{C}}\chi - (23.47 \pm 2.23) ^3\chi_{\text{c}} + (40.45 \pm 5.57) ^4\chi_{\text{c}} - (5.70 \pm 1.42) ^4\chi_{\text{pc}} + (10.91 \pm 1.81) ^4\chi_{\text{pc}}^{\text{v}} - (135.04 \pm 10.05)$$

USED SOFTWARE

The models based on topological descriptors were created with JBSMM (Java Based System for Molecular Modeling). It is an in-house developed software system [5] that supports the main stages of the molecular modeling: structure representation, descriptor calculation, MLR model creation and model statistics and validation. All quantum chemistry computations were carried out with the Gaussian 98W [A3] on level HF/6-31G(d) [6].

ACKNOWLEDGEMENTS

We would like to thank the Bulgarian National Fund for Scientific Research NFNI (project VUH-17/05) for supporting this scientific work.

REFERENCES

1. Smeeks, F.C. and Jurs, P.C. Prediction of boiling points of alcohols from molecular structure. *Anal. Chim. Acta.* 233, 111-119 (1990).
2. Katritzky, A.R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* 100, 10400-10407 (1996).

3. Chunhui Lu; Weimin Guo; Yang Wang; Chunsheng Yin. Novel distance-based atom-type topological indices DAI for QSPR/QSAR studies of alcohols. *J. Mol. Model.* 12, 749–756 (2006)
4. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH, 2000.
5. N. Kochev, O. Pukalov; Software system for molecular modeling – JBSMM; Balkan Conference of Young Scientists; 16-18 June, 2005, Plovdiv, Bulgaria.
6. Gaussian 98, Revision A.7, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.

