



Quantitative Structure-Properties Relationship Studies on Physicochemical Properties of Organic Molecules Using CODESSA†

HUI ZHANG¹, JUN LI² and CHAN KYUNG KIM^{2,*}

¹Department of Chemical and Materials Engineering, Hefei University, Hefei 230022, P.R. China

²Department of Chemistry, Inha University, 100 Inha-ro, Nam-gu, Incheon 402-751, Republic of Korea

*Corresponding author: E-mail: kckyung@inha.ac.kr

AJC-13300

Six physicochemical properties of organic molecules, normal boiling points, heats of vaporization, heats of sublimation, heats of fusion, liquid density and solid density, were predicted by quantitative structure-properties relationship (QSPR) approach. The molecules in each set were optimized using semi-empirical AM1 and PM3 hamiltonians and verified as minima from frequency calculations using the same levels of theory. CODESSA package was then used to calculate molecular descriptors and to perform linear regressions to find out the dual-parameter equations. The results of best correlations were similar to those published earlier. The method applied in this work can be extended to predict other physicochemical properties with confidence.

Key Words: Prediction of physicochemical properties, QSPR, Semi-empirical methods, CODESSA.

INTRODUCTION

In chemistry and material researches the synthesis of a new compound is an essential step in developing new materials with improved properties. Lots of synthetic works have been completed with success. In usual laboratory work, new target molecules are usually sketched and designed by trial-and-error method and the properties of successful compounds are measured later by use of instrumental analyses. In this process, many compounds that cannot satisfy the desired properties are abandoned. To save time and efforts, rational design is an essential process in chemical synthesis and drug design works. At the beginning, many researchers already knew the desired properties or the skeleton of their target molecules. Before synthesis, therefore, it is reasonable to predict the physicochemical properties using quantitative structure-properties relationship¹. Based on the prediction, the proposed molecules can be adopted or abandoned. In this sense, quantitative structure-properties relationship study is important for material/drug design processes. Quantitative structure-properties relationship is develop a function or a series of functions to estimate the properties of molecules. There are many approaches in quantitative structure-properties relationship study-fragment based method, neural network and multiple linear/nonlinear regression methods¹. The data in quantitative structure-properties relationship study is a set of molecules whose properties

were already determined by experimental methods. Then some variables called descriptors were calculated using various methods. The most sophisticated way of calculating molecular descriptors involved three-dimensional (3D) structures of molecules. 3D structures can be calculated with the help of molecular mechanics or quantum chemical packages. *Ab initio* and density functional theory can give accurate results at the expenses of excessive CPU time². These methods may experience some difficulty if the database under consideration contains a few hundred to a few thousand molecules. To balance between time and accuracy, some semi-empirical method³ such as AM1⁴ or PM3⁵ was chosen as an alternative. The physicochemical properties of molecules, denoted as object functions hereafter, are then predicted from the molecular descriptors using statistical procedure. In many quantitative structure-properties relationship studies, the object function is normally a linear combination of descriptors and the contributing descriptors are then selected based on their characteristics. Comprehensive descriptors for structural and statistical analysis (CODESSA)⁶ is a convenient and efficient package that can do both jobs of descriptor evaluation and statistical analysis. In this work, quantitative structure-properties relationship study was performed to derive some reasonable objective functions for six physicochemical properties of organic molecules. The results obtained in this work were compared with earlier publications.

†Presented to the 6th China-Korea International Conference on Multi-functional Materials and Application, 22-24 November 2012, Daejeon, Korea

EXPERIMENTAL

Calculation: The data set considered in this work is the same as that in earlier work⁷⁻¹⁰. The molecular structures were optimized using semi-empirical AM⁴ and PM3 methods⁵ and the geometries were verified as minima having all positive force constants by frequency calculations. The optimized structures corresponded to global minima or at least local minimum in the potential energy surface. All the structures were minimized using MOPAC 6.0¹¹. Then CODESSA package calculated the molecular descriptors using these optimized structures. This package calculated six different types of descriptors-constitutional, topological, geometrical, electrostatic, quantum-chemical and thermodynamic descriptors. The total number of descriptors depended on the molecular formula and structures but is approximately over 350 per molecule. Two types of statistical approaches such as the heuristic and best multi linear-regression were tried to obtain the best objective functions. In this work, objective functions with two descriptors were selected because two descriptors were enough to obtain satisfactory object functions in earlier works⁷⁻¹⁰. If the number of descriptors increases, the correlation may improve but statistical significance may reduce in case of small database.

RESULTS AND DISCUSSION

The six physicochemical properties of organic molecules, normal boiling points (T_{bp})^{7,10}, heats of vaporization (ΔH_{vap})^{7,10}, heats of sublimation (ΔH_{sub})^{8,10}, heats of fusion (ΔH_{fus})^{8,10}, liquid density^{9,10} and crystal density^{8,10} were predicted using CODESSA package. These properties were already predicted using the general interaction properties function (GIPF)⁷, which calculates electrostatic potential (ESP) on the molecular surface. The derived dual-parameter relationships are shown in eqns. (1)-(6).

$$T_{bp} = \alpha(\text{AREA}) + \beta\sqrt{v\sigma_{tot}^2} + \gamma \quad (1)$$

$$\Delta H_{vap} = \alpha\sqrt{\text{AREA}} + \beta\sqrt{v\sigma_{tot}^2} + \gamma \quad (2)$$

$$\Delta H_{sub} = \alpha(\text{AREA})^2 + \beta(v\sigma_{tot}^2) + \gamma \quad (3)$$

$$\Delta H_{fus} = \alpha(\text{AREA}) + \beta(v\pi) + \gamma \quad (4)$$

$$\text{liquid density} = \alpha\left(\frac{M}{\text{AREA}}\right) + \beta(\pi) + \gamma \quad (5)$$

$$\text{crystal density} = \alpha\left(\frac{M}{\text{AREA}}\right) + \beta\left(\frac{\sigma_{tot}^2}{\pi}\right) + \gamma \quad (6)$$

where, AREA and M are molecular surface and molecular weight, respectively and v , σ_{tot}^2 and π are descriptors derived from ESP values. The summary of multiple linear correlations of these properties are summarized in Table-1.

CODESSA package calculates various descriptors using AM1 and PM3 optimized structures and also performs statistical analysis by use of two different algorithms. The best dual-parameter equations (shown in eqn. 7) obtained for six physicochemical properties are summarized in Table-2.

TABLE-1
MULTIPLE LINEAR CORRELATIONS IN EQS. (1)–(6)

Physical property	N ^a	r ^b	
		vdW ^c	Density ^d
T_{bp}	100	0.936	0.948
ΔH_{vap}	41	0.947	0.965
ΔH_{sub}	34	0.951	0.950
ΔH_{fus}	37	0.928	0.919
Liquid density	61	0.984	0.982
Crystal density	36	0.988	0.987

^aNumber of molecules considered. ^bLinear regression coefficient.

^cFrom ref. 7. ^dFrom ref. 8

TABLE-2
DUAL-PARAMETER CORRELATIONS
OF SOME PHYSICO-CHEMICAL PROPERTIES

Physical property	Method	x_1^a	x_2^a	α^b	β^b	γ^b	r ^c
T_{bp}	AM1	S_1^d	S_2^e	8.050	20800	83.3	0.935
	PM3	S_2	S_3^f	3.970	20900	18.0	0.941
ΔH_{vap}	AM1	S_4^g	S_2	0.106	16.8	-3.59	0.922
	PM3	S_5^h	S_3	0.365	18.5	-3.69	0.943
ΔH_{sub}	AM1	S_5	S_6^i	3.340	0.809	1.00	0.969
	PM3	S_5	S_7^j	3.360	273.	0.869	0.953
ΔH_{fus}	AM1	S_8^k	S_9^l	0.106	16.8	-3.59	0.950
	PM3	S_1	S_9	0.365	18.5	-3.69	0.950
Liquid density	AM1	S_{10}^m	S_{11}^n	-0.00657	0.173	-5.36	0.957
	PM3	S_{10}	S_{11}	-0.00670	0.171	-5.29	0.957
Crystal density	AM1	S_{12}^o	S_{10}	-0.0512	0.163	-4.88	0.967
	PM3	S_{11}	S_{10}	-0.00644	0.198	-6.18	0.970

^aDescriptors in dual-parameter equations. ^bCoefficients of multiple correlation in eq. 7. ^cLinear regression coefficient. ^dXY shadow. ^eHACA-2/TMSA (Zefirov's PC). ^fALFA polarizability (DIP). ^gKier & Hall index (order 1). ^hRandic index (order 1). ⁱHDCA H-donors charged surface area (QM). ^jHA dependent HDCA-2/SQRT(TMSA). ^kMolecular surface area. ^lImage of the Onsager-Kirkwood solvation energy. ^mZero-point vibrational energy. ⁿTranslation entropy (300K). ^oNumber of H atoms

$$\text{Physical property} = \alpha(x_1) + \beta(x_2) + \gamma \quad (7)$$

Table-2 shows that the best correlations for six physical properties can be described by different types of descriptors except for liquid density. The magnitudes of coefficients, α and β , however, are comparable for AM1 and PM3 except for heats of sublimation. Inspection of Table-2 shows that the linear correlation coefficients (r in the last column) are almost same for both methods, AM1 and PM3, which suggests that the optimized geometries are quite similar for organic molecules considered in this work. Surprisingly, the linear correlation coefficients summarized in Table-2 are comparable to the corresponding coefficients in Table-1. To obtain the results in Table-1, molecular structures were optimized at the B3LYP/6-31G(d) level of theory and the ESP values were calculated on the envelop of electron density⁷ or on the van der Waals surface⁸ at the same level of theory and finally the dual-parameter object functions were selected using successive multiple linear regressions¹¹. Compared to such a high level quantitative structure-properties relationship study, the approach using semi-empirical and CODESSA packages is fast but still reliable to give quantitative results.

Conclusion

Six physicochemical properties of organic molecules, normal boiling points, heats of vaporization, heats of subli-

mation, heats of fusion, liquid density and crystal density, were examined in this work. The combination of semi-empirical method and CODESSA package can give satisfactory results compared to those obtained using a high-level density functional theory method, B3LYP/6-31G(d). This work suggests that the physicochemical properties of organic molecules can be predicted using the approach applied in this work.

ACKNOWLEDGEMENTS

This work was supported by Inha University and Anhui Provincial Natural Science Foundation (1308085QB30).

REFERENCES

1. P.C. Jurs, in ed.: P.V.R. Schleyer, Encyclopedia of Computational Chemistry; John Wiley & Sons, New York, Vol. 4, p. 2320 (1998).
2. W.J. Hehre, L. Radom, P.V.R. Schleyer and J.A. Pople, Molecular Orbital Theory, John Wiley & Sons; New York (1986).
3. J. Sadlej, Semi-Empirical Methods of Quantum Chemistry, Ellis Horwood Ltd., Chichester (1985).
4. M.J.S. Dewar, E.G. Zoebisch and E.F. Healy, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
5. J.J.P. Stewart, *J. Comput. Chem.*, **10**, 209 (1989).
6. CODESSA 2.7.16 was available from Semichem, Inc. (2012).
7. P. Politzer and J.S. Murray, Quantitative Treatments of Solute/Solvent Interactions, Elsevier, Amsterdam, p. 243 (1994).
8. P. Politzer, J.S. Murray, M.E. Grice, M. Desalvo and E. Miller, *Mol. Phys.*, **91**, 923 (1997).
9. J.S. Murray, T. Brince and P. Politzer, *Chem. Phys.*, **204**, 289 (1996).
10. C.K. Kim, K.A. Lee, K.H. Hyun, H.J. Park, I.Y. Kwack, C.K. Kim, H.W. Lee and B.-S. Lee, *J. Comput. Chem.*, **25**, 2073 (2004).
11. MOPAC 6.0, J.J.P. Stewart, Frank J. Seiler Research Laboratory, U.S. Air Force Academy, Colorado Springs, CO 80840.