

UDC 541.65

ESTIMATION AND PREDICTION OF ^{13}C NMR CHEMICAL SHIFTS OF CARBON ATOMS IN BOTH ALCOHOLS AND THIOLS**J. Tong^{1,2}, Y. Chen^{1,2}, S. Liu^{1,2}, X. Xu^{1,2}, F. Cheng^{1,2}**

¹College of Chemistry and Chemical Engineering, Shaanxi University of Science and Technology, Xi'an, PR China, e-mail address: Jianbotong@yahoo.com.cn

²Key laboratory of Auxiliary Chemistry & Technology for Chemical Industry, Ministry of Education. Shaanxi University of Science and Technology, Xi'an, PR China

Received July, 2, 2011

Quantitative structure-spectrum relationship calculations of ^{13}C NMR chemical shifts of both 302 carbon atoms in 56 alcohols and 62 carbon atoms in 15 thiols are described using several parameters: the atomic ionicity index (INI), the polarizability effect index (PEI), and stereoscopic effect parameters (ϵ) of the compounds. The ^{13}C NMR chemical shifts for these compounds of both alcohols and thiols can be estimated through the multiple linear regression (MLR). A MLR model was built with variable screening by the stepwise multiple regression and examined by validation on its stability. The correlation coefficient of the established model as well as the leave-one-out cross-validation was 0.9724 and 0.9716 respectively. The results obviously indicate that INI and ϵ are linearly related with ^{13}C NMR chemical shifts, which provides a new method for calculating ^{13}C NMR chemical shifts in the compounds of both alcohols and thiols.

Keywords: quantitative structure-spectrum relationship (QSSR), ^{13}C NMR chemical shifts of alcohols and thiols, atomic ionicity index (INI), polarizability effect index (PEI), stereoscopic effect parameters (ϵ).

INTRODUCTION

As is known, nuclear magnetic resonance (NMR) spectroscopy is a useful tool for the structure elucidation of compounds [1], especially ^{13}C NMR spectroscopy plays an important role in investigating the structures of many organic compounds in chemistry, biology, medicine, et al. [2, 3]. Studies on ^{13}C NMR spectroscopy have attracted more and more interest of scientists. Grant and Paul [4] and Lindeman and Adams [5] first proposed a ^{13}C NMR simulation for alkanes and later for compounds containing heteroatoms and/or rings. Recently, a ^{13}C NMR simulation has been developed due to the advancement of computer science, computational chemistry [6—9], calculation methodology such as *ab initio* and density functional calculations [10—14], automatic structure identification [15, 16], graph theoretical techniques and topological indices [17, 18]. Recently, Jaiswal and Khadikar [19] have calculated ^{13}C NMR chemical shifts on 32 carbinol carbon atoms with Wiener (W), Padmakar-Ivan (PI), and Connectivity ($^m\chi$, $^m\chi^v$) indices and got a good result. Furthermore, they [20] used them as a molecular descriptor for modeling the lipophilicity (logP) of alcohols, while Liu [17] used the atomic electronegative distance vector (AEDV) to express the ^{13}C NMR chemical shifts of both alcohols and alkanes.

In this paper, the atomic ionicity index (INI) [21], polarizability effect index (PEI) [22], and stereoscopic effect parameters (ϵ) were first proposed for predicting ^{13}C NMR chemical shifts of alcohols and thiols. The developed model was simple, convenient, and highly accurate. Moreover, each

item of the model had very explicit physical significance. The obtained multiple linear regression (MLR) equation not only clarified the relationship between the ^{13}C NMR chemical shifts of alcohols, thiols, and their molecular structures, but also gave theoretical laws for predicting ^{13}C NMR chemical shifts and studying the relationship between the structure and property for alcohols and thiols.

THEORY AND CALCULATION

As we know, alcohols and thiols are considered as derivatives of alkanes, while alkanes are the derivatives of methane CH_4 . According to the principle of NMR spectroscopy, isotopic atomic nuclei in a molecule are differently shielded by their electron clouds due to their different chemical environments. As a result, different resonance frequencies are observed during an NMR experiment exciting these isotopes. If these frequencies are measured as differences to the resonance frequency of the inner standard, they are designated as chemical shifts.

The energy of the spin of the carbon nucleus in CH_4 is expressed as: $\Delta E^0 = h\nu_0$; as expected, the energy of the carbon nuclear spin when $-\text{OH}$ or $-\text{SH}$ substitute for H in CH_4 is $\Delta E = h\nu$. Therefore, $-\text{OH}$ or $-\text{SH}$ substitution for H causes a change in energy: $\Delta\Delta E = h\Delta\nu_0$, thus

$$\nu = \nu_0 + \Delta\nu_0. \quad (1)$$

In Eq. 1, $\Delta\nu_0 = \Delta\Delta E/h$, which is directly related to the polarizability effect of alkyl-substituents, is the change of electronegativity of the central carbon atom and the stereoscopic effect on the central carbon atom.

Using PEI, we denote the polarizability effect index; $\sum\text{PEI}$ expresses the sum of the polarizability effect indices of alkyl groups connected to the central carbon atom. Besides, the electric field of the central carbon atom is related to the electronegativity of polarity groups, which has direct relevance to INI. In addition, the stereoscopic effect on the central carbon atom is ε . Therefore,

$$\nu = A + B \cdot \sum\text{PEI} + C \cdot \text{INI} + D \cdot \varepsilon, \quad (2)$$

where PEI is polarizability effect index; INI is atomic ion index; ε is spectroscopic effect parameter.

Carbon atoms of alcohol and thiol molecules have a special tetrahedral structure. It is well known that the presence of attached and nearby atoms has a profound effect on the ^{13}C NMR chemical shifts of the central carbon atom. The effect of substituents on ^{13}C NMR chemical shifts can be felt from as far as four bonds away or directly through the space [23]. In this paper, we calculated the effect of substituents on ^{13}C NMR chemical shifts as far as six bonds away from the central carbon atom, and the results showed that atoms or groups which apart from three bonds away carbon nucleus have slight effect on the ^{13}C NMR chemical shifts. According to the aforementioned results and discussion, we used stereoscopic effect parameters to predict the ^{13}C NMR chemical shifts. The stereoscopic effect parameters were named as ε_i ($i = \alpha, \beta, \gamma$)

$$\varepsilon_i = n_{iC} + n_{iO} \times q_{O/C} \quad \text{or} \quad \varepsilon_i = n_{iC} + n_{iS} \times q_{S/C}. \quad (3)$$

In above Eq. 3, n_{iC} is the number of carbon atoms on i -substituents; n_{iO} and n_{iS} are the number of oxygen and sulfur atoms on i -substituents, and $q_{O/C}$, $q_{S/C}$ are the relative electronegativity of oxygen and sulfur atoms, i.e. the ratio of oxygen or sulfur Pauling electronegativity to that of the carbon atom; namely, $q_{O/C} = 3.44/2.55$, $q_{S/C} = 2.58/2.55$.

The aforementioned discussion and results prompted us that INI, PEI, and ε_i can characterize the molecular property and that they can be used as molecular descriptors for modeling ^{13}C NMR chemical shifts.

RESULTS AND DISCUSSION

MLR model. The dataset was composed of 302 carbon atoms in 56 alcohols and 62 carbon atoms in 15 thiols. Their parameters INI, $\sum\text{PEI}$, and ε_i are given in [24, 25]. MLR is a classical modeling method; it can obtain an optimal result in least squares (LS) significance by linear fitting independent variables and dependent variables. MLR models were established to study relationships between INI, PEI, ε_i and ^{13}C NMR chemical shifts of 302 carbon atoms in 56 alcohols, 62 carbon atoms in 15 thiols, and 364 carbon atoms in both 56 alcohols and 15 thiols respectively.

$$CS = 92.9498 - 9.8503INI + 15.9796\Sigma PEI - 8.1140\varepsilon_{\alpha} + 1.3566\varepsilon_{\beta} - 2.2933\varepsilon_{\gamma} \quad (4)$$

$$n = 302; R = 0.9819; SD = 3.7429; F = 1592.2881$$

$$CV: n = 302; R_{CV} = 0.9804; SD_{CV} = 3.8932; F_{CV} = 1467.1809$$

$$CS = 35.9801 - 3.2946INI + 0.2095\Sigma PEI + 6.1042\varepsilon_{\alpha} + 7.7329\varepsilon_{\beta} - 2.5674\varepsilon_{\gamma} \quad (5)$$

$$n = 62; R = 0.9839; SD = 1.6286; F = 340.3943$$

$$CV: n = 62; R_{CV} = 0.9782; SD_{CV} = 1.8976; F_{CV} = 247.7826$$

$$CS = 89.3086 - 9.4138INI + 15.6212\Sigma PEI - 7.4535\varepsilon_{\alpha} + 1.6715\varepsilon_{\beta} - 2.7131\varepsilon_{\gamma} \quad (6)$$

$$n = 364; R = 0.9764; SD = 3.9997; F = 1460.6350$$

$$CV: n = 364; R_{CV} = 0.9749; SD_{CV} = 4.1155; F_{CV} = 1375.5869.$$

In Eqs. (4)–(6), n is the number of samples used for model building, R is the multiple correlation coefficient, SD is the standard error of estimation, F is Fisher's statistics, CV means leave-one-out cross validation.

In order to avoid casual correlation, SMR was employed for variable screening. Firstly, a proper F value was fixed so that each variable was introduced into the equations gradually according to importance; then we observed a change in a series of parameters with an increase in variables. It was indicated that with the increase in the number of variables (m), the values of the correlation coefficient (R) and R_{CV} of leave-one-out cross validation increase gradually. However, a change in SD and SD_{CV} showed an approximately reverse process; the values of SD and SD_{CV} of the MLR models decreased with an increase in variables. It was found that when m was 5, SD_{CV} reached the least value, R took the largest value, and R_{CV} was relatively high. However, the t -test of Eqs. (4)–(6) showed that 4 variables (INI, ε_{α} , ε_{β} , and ε_{γ}) were distinctly significant (the minimum $|t|$ was 2.752 larger than critical value 2) and did not have multicollinearity (the highest variable inflation factor (VIF) was 7.809, which was less than the experimental value of 50). Moreover, the inter-correlation coefficient between all pairs of descriptors and dependent variable chemical shifts was listed in Table 1.

It could be seen that the correlation coefficients between PEI and INI, and PEI and ε_{α} in the set of 56 alcohols and 15 thiols (Eq. 6) were high: up to 0.8035 and 0.9545. Therefore, the 4-variable models have the best prediction capability and stability. MLR models (4) were established to study relationships between INI, ε_{α} , ε_{β} , ε_{γ} and ^{13}C NMR chemical shifts of 364 samples. It was presented as follows:

$$CS = 112.8600 - 9.6331INI - 1.1724\varepsilon_{\alpha} + 2.6094\varepsilon_{\beta} - 2.5475\varepsilon_{\gamma} \quad (7)$$

$$n = 364; R = 0.9724; SD = 4.3089; F = 1560.5120$$

$$CV: n = 364; R_{CV} = 0.9716; SD_{CV} = 4.3693; F_{CV} = 1515.2219.$$

The negative INI coefficients in above Eq. (7) indicated that the ^{13}C NMR chemical shifts were directly (linearly) related to the magnitude of INI, and the results gave rise to the higher field resonance. In addition, the presence of attached and nearby atoms had a profound effect on the ^{13}C NMR chemical shifts, i.e., atoms in different positions away from the central carbon atom gave rise to the shielding or deshielding effect on the carbon nucleus. What is more, the magnitude of coefficients in Eq. (7) reflected that the descriptors had different contributions to the ^{13}C NMR chemical shifts.

Table 1

Covariance matrix of structural parameters together with chemical shifts

Parameters	INI	PEI	ε_{α}	ε_{β}	ε_{γ}	$Y(CS)$	Average	SD
INI	1.0000	-0.8035	-0.8015	-0.2080	0.2772	-0.9537	8.2576	1.7829
PEI	-0.8035	1.0000	0.9545	0.2547	-0.2791	0.7902	2.3202	0.4045
ε_{α}	-0.8015	0.9545	1.0000	0.1113	-0.3054	0.7471	1.9800	0.9019
ε_{β}	-0.2080	0.2547	0.1113	1.0000	-0.0893	0.3379	1.9550	0.9726
ε_{γ}	0.2772	-0.2791	-0.3054	-0.0893	1.0000	-0.3850	1.4169	0.9462
$Y(CS)$	-0.9537	0.7902	0.7471	0.3379	-0.3850	1.0000	32.4847	18.3746

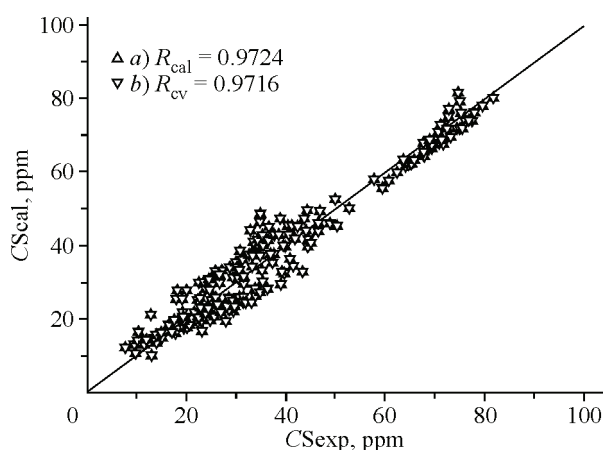


Fig. 1. Plot of the experimented vs. calculated chemical shifts of 302 carbon atoms in 56 alcohols and 62 carbon atoms in 15 thiols for Cal (a), CV (b)

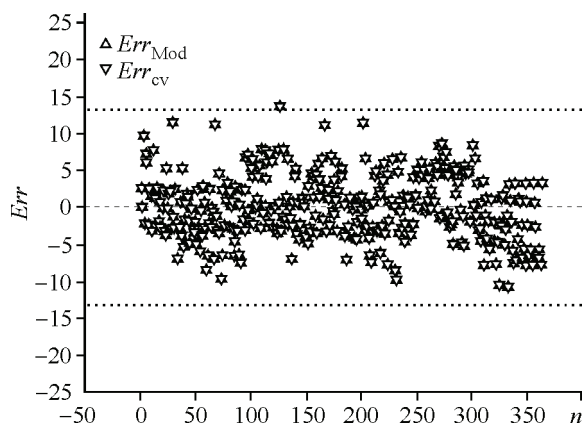


Fig. 2. Residual distribution of the built model; cross validation (LOO) for 364 samples. (The dashed line is three-standard-deviation errors)

From the data [24, 25], one can know that better agreement between the experimental and estimated data indicates that the INI and ε_i descriptors proposed in our calculations were correct and effective.

Model validity. An excellent model should not only have good estimated results of internal samples, but also high predictive ability. Eq. (7) was used to predict 364 samples. The estimated results related to the experimental results are presented in Fig. 1 and the residual distribution is shown in Fig. 2.

It was indicated that INI, ε_α , ε_β , and ε_γ were obviously related to the ^{13}C NMR chemical shifts of alcohols and thiols, because in Fig. 1, all samples were dispersed equally along the one-to-one correlation line through the origin, forming an angle of $\sim 45^\circ$; and in Fig. 2, most of the residual distribution was within triple standard deviations (i.e. within the area enclosed by a dotted line). The similar conclusion was obtained when comparing the Cook distance of samples with the centered leverage values (Fig. 3); only samples 1, 29, 127, and 321 were slightly extraordinary, which might be attributed to different experimental conditions or insufficiency of INI, ε_α , ε_β , and ε_γ in the structure feature characterization, or insufficiency of the model in establishing a link between the structural parameters and the property, or particularity of the samples themselves.

It was found in Fig. 3 that sample 1 was a carbon atom in methanol, which was connected with 3 H atoms. Samples 29, 127, and 321 were quaternary carbon atoms with 4 carbon atoms around them. Since they had an exceptional structure, their deviations were greater. Considering that they were not exceptional too much, they were left in eq. (7).

An excellent QSAR/QSPR/QSSR model should have not only good estimation stability for any internal sample, but also good predictive ability for an external sample. Although it has been proved that the above models have good estimation abilities for the ^{13}C NMR chemical shifts of alcohols and thiols, the prediction abilities have to be further tested for external samples. The most usual method to test the predictive ability is the CV method with the leave-one-out procedure. Here each time one of the samples is removed from the set, and a model is developed with the remaining $n-1$ samples. At each step, the ^{13}C NMR chemical shift of the removed sample is predicted by the obtained model. Another cross-validation parameter of interest is the predictive correlation coefficient that is usually named as the cross-validation R_{CV}^2 . The highest value of R_{CV}^2 indicates the best predictive power. Once again, R_{CV}^2 (0.9414) were in favor of eq. (7). A comparison of the calculated

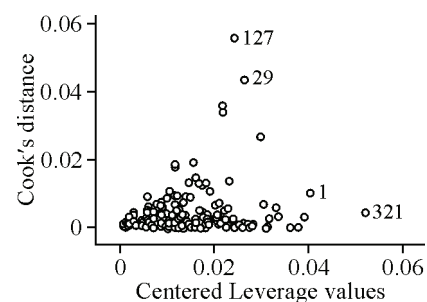


Fig. 3. Cook's distance vs. the centered leverage values of model (4)

or predicted ^{13}C NMR chemical shifts and the experimental data for 56 alcohols and 15 thiols is shown in Fig. 1. One can well find that the calculated, either estimated or predicted, ^{13}C NMR chemical shifts values are in good agreement with the experimental data.

CONCLUSIONS

It can be seen within Eq. (7) that the correlations of the ^{13}C NMR chemical shifts in alcohols and thiols are excellent. The correlation coefficient is $R = 0.9724$, and the expressions are suitable to calculate the ^{13}C NMR chemical shifts of alcohols and thiols. In addition, Eq. (7) describes a new, simple, efficient method to calculate ^{13}C NMR chemical shifts in compounds of either alcohols or thiols. From the aforementioned results it can be concluded that INI in combination with ε_i can be used successfully for modeling ^{13}C NMR chemical shifts. The obtained results show that ^{13}C NMR chemical shifts can be predicted based on the molecular structure alone. It is foreseeable that INI and ε_i would be utilized widely in QSSR and/or other related studies. Similar studies of the ^{13}C NMR chemical shifts of aliphatic compounds, such as amine, ether compounds, are in progress.

Acknowledgments. This work was supported by the National Natural Science Funds (21275094), the Scientific Research Planning Program of Shaanxi Province of China (2011K07-13), the Scientific Research Planning Program of the Education Department of Shaanxi Province (12JK0629) (11JK0602), the Scientific Research Planning Program of Key Laboratory of Shaanxi Province of China (11JS022), the Scientific Research Planning Program of Xianyang (2011K09-09), the Scientific Research Planning Program of Yulin City and the Graduate Innovation Fund of Shaanxi University of Science and Technology.

REFERENCES

1. Benzi C., Cossi M., Barone V., Tarroni R., Zannoni C.A. // *J. Phys. Chem. B.* – 2005. – **109**. – P. 2584 – 2590.
2. Liu X.H., Ren Y.R., Zhou P., Shang Z.C. // *J. Mol. Struct.* – 2011. – **995**, N 1-3. – P. 163 – 172.
3. Neuvonen H., Neuvonen K. // *J. Chem. Soc., Perkin Trans. 2.* – 1999. – P. 1497 – 1502.
4. Grant D.M., Paul E.G. // *J. Amer. Chem. Soc.* – 1964. – **86**. – P. 2984 – 2990.
5. Lindeman L.P., Adams J.Q. // *Anal. Chem.* – 1971. – **43**, N 10. – P. 1245 – 1252.
6. Clouser D.L., Jurs P.C. // *J. Chem. Inf. Comput. Sci.* – 1996. – **36**. – P. 168 – 172.
7. Ivanciuc O., Rabine J.P., Cabrol-Bass D. // *Comput. Chem.* – 1997. – **21**, N 6. – P. 437 – 443.
8. Ivanciuc O., Rabine J.P., Cabrol-Bass D., Panaye A., Doucet J.P. // *J. Chem. Inf. Comput. Sci.* – 1997. – **37**. – P. 587 – 598.
9. Nohair M., Zakarya D. // *J. Chem. Inf. Comput. Sci.* – 2002. – **42**. – P. 586 – 591.
10. Meng Z.Z., Carper W.R. // *J. Mol. Struct. (Theochem).* – 2002. – **588**. – P. 45 – 53.
11. Crescenzi O., Correale G., Bolognese A., Piscopo V., Parrilli M., Barone V. // *Org. Biomol. Chem.* – 2004. – **2**. – P. 1577 – 1581.
12. Sun G.Y. // *Chem. Phys.* – 2003. – 289. – P. 371 – 380.
13. Balandina A., Mamedov V., Franck X., Figadère B., Latypov S. // *Tetrahedron Lett.* – 2004. – **45**. – P. 4003 – 4007.
14. Moss W.N., Goroff N.S. // *J. Org. Chem.* – 2005. – **70**. – P. 802 – 808.
15. Militão J.S.L.T., Emerenciano V.P., Ferreira M.J.P., Cabrol-Bass D., Rouillard M. // *Chemometr. Intell. Lab. Systems.* – 2003. – **67**. – P. 5 – 20.
16. Satoh H., Koshino H., Uzawa J., Nakata T. // *Tetrahedron.* – 2003. – **59**. – P. 4539 – 4547.
17. Liu S.S., Xia Z.N., Liu Y., Cai S.X., Li Z. // *Chin. J. Chem.* – 2000. – **18**, N 2. – P. 165 – 174.
18. Panaye A., Doucet J.P., Fan B.T. // *J. Chem. Inf. Comput. Sci.* – 1993. – **33**. – P. 258 – 265.
19. Jaiswal M., Khadikar P. // *Bioorgan. & Medicin. Chem. Lett.* – 2004. – **12**. – P. 1793 – 1798.
20. Khadikar P.V., Sharma V., Varma R.G. // *Bioorgan. & Medicin. Chem. Lett.* – 2005. – **15**. – P. 421 – 425.
21. Nie C.M., Li Z.H., Wen S.N. // *Chin. J. Organ. Chem.* – 2002. – **22**, N 1. – P. 46 – 51. (in Chinese)
22. Cao C.Z., Li Z. // *J. Chem. Inf. Comput. Sci.* – 1998. – **38**. – P. 1 – 7.
23. Beger R.D., Wilkes J.G. // *J. Chem. Inf. Comput. Sci.* – 2001. – **41**. – P. 1322 – 1329.
24. *Sadtler Standard Carbon-13 NMR Spectra.* Sadtler Research Laboratories Division of Bio-Rad Laboratories, INC. Printed in USA, 1980.
25. http://www.aist.go.jp/RIODB/SDBS/cgi-bin/cre_index.cgi.