An atomic electronegative distance vector and carbon-13 nuclear magnetic resonance chemical shifts of alcohols and alkanes

LIU, Shu-She[a,b](刘树深) XIA, Zhi-Ning[c](夏之宁) CAI, Shao-Xi[a](蔡绍皙) LIU, Yan[a](刘堰)
LI, Zhi-Liang*[,a,c](李志良)

[a] *College of Bioengineering, Chongqing University, Chongqing 400044, China*
[b] *Department of Applied Chemistry, Guilin Institute of Technology, Guilin, Guangxi 541004, China*
[c] *College of Environmental and Chemical Engineering, Chongqing University, Chongqing 400044, China*

A novel atomic electronegative distance vector (AEDV) has been developed to express the chemical environment of various chemically equivalent carbon atoms in alcohols and alkanes. Combining AEDV and γ parameter, four five-parameter linear relationship equations of chemical shift for four types of carbon atoms are created by using multiple linear regression. Correlation coefficients are $R = 0.9887$, 0.9972, 0.9978 and 0.9968 and roots of mean square error are RMS = 0.906, 0.821, 1.091 and 1.091 of four types of carbons, *i.e.*, type 1, 2, 3, and 4 for primary, secondary, tertiary, and quaternary carbons, respectively. The stability and prediction capacity for external samples of four models have been tested by cross-validation.

Keywords Atomic electronegative distance vector (AEDV), γ calibration, simulation of carbon-13 nuclear magnetic resonance spectroscopy, alcohols and alkanes, molecular modelling

Introduction

Carbon-13 nuclear magnetic resonance (^{13}C NMR) spectrum simulation involves converting structural information into a simulated spectrum. Spectral simulation can be very useful in aiding chemists in the solution of complex structure elucidation problems and in the verification of chemical shift assignments. The three most common methods of ^{13}C NMR spectral simulation involve linear additivity relationships, database retrieval techniques, and empirical modeling.[1] To realize the computer simulation and the automatic elucidation of a spectrum, the investigations on the correlation between the structure and its spectrum have to be done. One of the most important procedures among various relationship researches is chemical structural parameterization or selection of descriptor variables. A number of empirical parameters had been developed and recently a lot of attentions have been paid on employment of more general structural parameters, in particular, those derived from chemical graph theory.[2-11] ^{13}C NMR spectral simulation started with the studying on alkanes by some authors, Grant and Paul,[12] Lindeman and Adams.[13] Afterwards, by using multiple linear regression, a modified simulation method was developed and the scope of application was expanded by Small and Jurs.[14,15] Recently, many chemometric methods predicting ^{13}C NMR chemical shift of organic compounds have been developed by means of artificial neural network algorithm[16-19] or multiple linear regression method.[20]

It has been found that the ^{13}C NMR chemical shift (*CS*) of resonance atom in an organic compound is related not only to its hybrid type but also to its chemical environment, *i.e.* the distribution of electronic clouds around it. So, it is essential to propose a topological index to represent the electronic environment of atom. In this paper, we develop a novel atomic electronegative distance vector (AEDV) on the basis of the atomic distance-edge (ADE) vector in our previous paper.[21] This AEDV vector will be used to illustrate the chemical environment of the examined atom in the alkane and alcohol molecules by employing electronegative correction and

Received June 3, 1999; accepted September 22, 1999.
Project supported by Ministry of Education "Chunhui Program" and Fok Ying Tung Education Foundation.

distance correction. And four five-parameter equations between the chemical shifts of carbon-13 NMR spectroscopy and the AEDV and the γ descriptor of various chemically equivalent carbon atoms in the alkanes and alcohols have been developed by multiple linear regression (MLR). These models are used to estimate and predict the ^{13}C NMR chemical shift of various atoms in the two types of compounds examined with good results.

Principle and algorithm

Atomic electronegative distance vector (AEDV)

The ^{13}C NMR chemical shift (*CS*) of resonance atom in an organic compound is closely related not only to its hybrid orbit type but also to non-hydrogen atom number around it. All carbon atoms in the examined alcohols and alkanes can be classified as four types of atoms based on the number of C—C or C—O bonds connecting with the other carbon or oxygen atoms in the molecule. The four types are called as C1, C2, C3 and C4 for primary (CH_3-), secondary ($CH_2<$), tertiary (-CH<), and quaternary (>C<) carbons, respectively. Various oxygen atoms in alcohols are considered as pseudo-carbon atoms. The chemical shift of chemically equivalent carbon is also determined by the distribution of electronic clouds around it and the density of electronic clouds is a function of the electronegative of connecting atom and distance. So, atomic electronegative distance vector (AEDV) is defined to express the chemical environment of atoms in alcohol and alkane. If the relative electronegative (REN) and relative bond length (RBL) are defined in the condition of REN of carbon being equal to 1 and RBL of C—C being equal to 1, an AEDV is defined as follows:

$$a_{ik} = \sum_{j\neq k}^{\text{all}(j)} \frac{q_j}{d_{ij}^6} \qquad (k = 1,2,3,4) \qquad (1)$$

where q_j is the REN of the jth non-hydrogen atom connected to the ith equivalent carbon, k non-hydrogen atom type and d_{ij} a RBL of distance between the jth non-hydrogen and ith equivalent atom. The REN of carbon is 1 and the REN of oxygen is 3.44/2.55 = 1.3490.[22] The RBL of C—C bondis 1 and the RBL of C—O bond is 0.143 nm/0.154 nm = 0.9286.[22] From Eq. (1), there are 4 elements in the AEDV.

γ-Effect

The γ-position carbon atom which is apart three C—C bonds from the centered carbon has, in general, a negative effect on the *CS* value of examined central atom and which is called γ-effect. In order to eliminate this effect, a γ parameter is introduced as follows:

$$\gamma_i = n_{iC} + n_{iO} \times 1.3490 \qquad (2)$$

where n_{iC} and n_{iO} are atom numbers of carbon and oxygen located in γ-position from the ith equivalent carbon.

Example of calculating AEDV *and γ-parameter*

Taking 3, 3-dimethyl-2-butanol and 2, 2, 4-trimethylpentane as example molecules (see Fig. 1), the calculation procedure of atomic electronegative distance vector (AEDV) and γ parameter is explained as follows.

There are 4 chemically nonequivalent carbons (i = 1,2,3,4), noted in subscript in Fig. 1, in the 3,3-dimethyl-2-butanol molecule. And there are 7 non-hydrogen atoms (j = 1,2,3,4,5,6,7), noted in superscript. The atom types of 7 atoms are k = 1,3,4,1,1,1 and 1, respectively. In according to Eqs. (1) and (2), four entries of the AEDV and the γ parameter for the first chemically equivalent carbon can be calculated as follows:

$$a_{11} = \frac{1.349}{1.9286^6} + \frac{1}{3^6} + \frac{1}{3^6} + \frac{1}{3^6} = 0.03030$$

$$a_{12} = 0.0000$$

$$a_{13} = \frac{1}{1^6} = 1.0000 \qquad a_{14} = \frac{1}{2^6} = 0.0156 \qquad \gamma = 3$$

The AEDV and the γ parameter for other 3 equivalent carbon atoms of 3,3-dimethyl-2-butanol are calculated in the same method as above. The AEDV and the γ parameter for 4 equivalent carbons are listed in Table 1.

There are 5 chemically nonequivalent carbons (i = 1, 2, 3, 4, 5), noted in subscript, in the 2, 2, 4-trimethylpentane molecule. And there are 8 non-hydrogen atoms (j = 1,2,3,4,5,6,7,8), noted in superscript. The atom types of 8 atoms are k = 1,4,2,3,1, 1,1 and 1, respectively. In according to Eqs. (1) and (2), four entries of the AEDV and the γ parameter for

the first chemically equivalentcarbon can be calculated as follows:

$$a_{11} = \frac{1}{2^6} + \frac{1}{2^6} + \frac{1}{2^6} + \frac{1}{2^6} = 0.0317 \qquad a_{12} = \frac{1}{2^2}0.0156$$

$$a_{13} = \frac{1}{3^6} = 1.0014 \qquad a_{14} = \frac{1}{1^6} = 1.0000 \qquad \gamma = 1$$

The AEDV and the γ parameter for other 4 equivalent carbon atoms of the 2,2,4-trimethylpentane molecule are calculated in the same as the method above. The AEDV and the γ parameter for 5 equivalent carbons are also listed in Table 1.

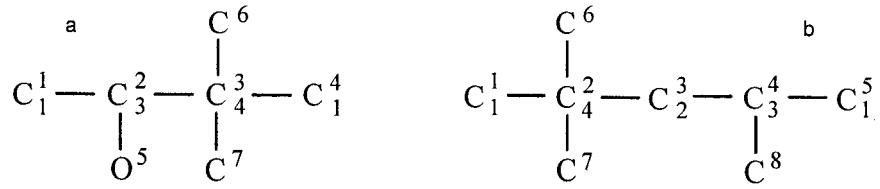

Fig. 1 A graph representing the carbon skeleton of 3,3-dimethyl-2-butanol (a) and 2,2,4-trimethylpentane (b).

Table 1 AEDV and γ-parameters for 9 chemically equivalent atoms of two sample compounds

No.	Compounds	i	Na	At	a_{i1}	a_{i2}	a_{i3}	a_{i4}	γ
1	C1-C2(OH)-C3(C,C)-C4	1	180	1	0.0303	0	1.0000	0.0156	3.0000
		2	181	3	3.1513	0	0	1.0000	0
		3	182	4	3.0418	0	1.0000	0	0
		4	183	1	0.0348	0	0.0156	1.0000	2.3490
2	C1-C2(C,C)-C3-C4(C)-C5	1	332	1	0.0317	0.0156	0.0014	1.0000	1
		2	333	4	3.0027	1.0000	0.0156	0	2
		3	334	2	0.0781	0	1.0000	1.0000	0
		4	335	3	2.0041	1.0000	0	0.0156	3
		5	336	1	0.0164	0.0156	1.0000	0.0014	1

Five-parameter equation of ^{13}C NMR *chemical shift*

For each of four types of carbon atoms, a simple relationship between the AEDV together with the γ parameter and chemical shift (CS) for carbon-13 NMR spectroscopy can be represented by:

$$CS_i = b_0 + \sum_{k=1}^{4} b_k \cdot a_{ik} + b_5 \cdot \gamma_i \qquad (3)$$

where b_0 and b_1, b_2, b_3, b_4, b_5 are intercept of regression equation and the regression coefficients, which are obtained by multiple linear regression method (MLR). The stability and prediction capacity for the external samples are tested by cross-validation technique.

Results and discussion

Atom type

The NMR experiments show that chemical shift of equivalent carbon atom is closely related to its hybrid type. When the carbon atom belongs to sp^3, sp^2, or sp hybrid, chemical shift is significantly different. For example, the CS for atom located in the end of alkane is often less than 30 and the CS for atom belonging to the C = C is more than 100. So, atoms in alcohols and alkanes are classified as three types, *i. e.*, sp^3, sp^2, and sp type. On the other hand, the CS value for the same hybrid type varies as the number of non-hydrogen atoms connected to. Because non-hydrogen atoms can be classified as four types, *i. e.* types 1,2,3,4 and chemical shifts are different from each other, the CS equation for every atom type must be constructed. A chemical environment of various atoms for the same atom type is determined by the structure of the molecule, and interaction between atoms in the molecule is affected by electronegative and distance between atoms, and the interaction can be described by AEDV.

Data set

The 516 nonequivalent carbon atoms in 63 alkanes and 36 alcohols are selected to study the relationship between chemical shifts and AEDV and γ parameters. The

series number (No.) of 99 compounds, the coding number of equivalent carbon atoms in every compound, and atomic series number (Na) for all 516 atoms are listed in Table 2. The experimentally measured carbon-13 NMR chemical shifts (CS_{exp}) of 330 chemically nonequivalent carbon atoms from 63 alkane compounds with the number of carbon atoms per alkane spanned from three through tencarbons are taken from references.[12,13] If there are two or more measured chemical shift values for the same equivalent atom, their average value will be taken as a new one. The experimentally measured carbon-13 NMR chemical shifts (CS_{exp}) of 186 chemically nonequivalent carbon atoms from 36 alcohol compounds are taken from Ref. 23. Because the *CS* values in Ref. 23 was determined under the condition of the *CS* of CS_2 setting to be zero, the *CS* relative to tetramethyl silane (TMS), CS_{TMS}, are calculated by equation $CS_{TMS} = 192.5 - CS$. These 63 alkanes and 36 alcohols compose of a working data set (M0). These 516 nonequivalent atoms are classified as 4 types: C1, C2, C3 and C4 and nonequivalent atom numbers for 4 types of atoms are 182, 223, 78 and 33 atoms, respectively.

Table 2 516 equivalent carbon atoms for 99 alcohols and alkanes

No.	Compounds	Na	No.	Compounds	Na	No.	Compounds	Na
1	C1-C2-OH	1-2	34	HO-C1-C2-C3(C,C)-C4	176-179	67	C1-C2-C3(C-C)-C4-C5-C6	320-325
2	C1-C2-C3-OH	3-5	35	C1-C2(OH)-C3(C,C)-C4	180-183	68	C1-C2(C,C)-C3(C6)-C4-C5	326-331
3	C1-C2-C3-C4-OH	6-9	36	C1-C2(C,C)-C3(OH)-C(C,C)-C	184-186	69	C1-C2(C,C)-C3-C4(C)-C5	332-336
4	C1-C2-C3-C4-C5-OH	10-14	37	C1-C2-C	187-188	70	C1-C2(C)-C3(C6,C)-C4-C5	337-342
5	C1-C2-C3-C4-C5-C6-OH	15-20	38	C1-C2-C-C	189-190	71	C1-C2(C)-C3(C4)-C(C)-C	343-346
6	C1-C2(OH)-C	21-22	39	C1-C2(C)-C	191-192	72	C1-C2(C)-C3(C-C)-C4-C5	347-351
7	C1-C2(OH)-C3-C4	23-26	40	C1-C2-C3-C-C	193-195	73	C1-C2-C3(C-C,C4)-C-C	352-355
8	C1-C2(OH)-C3-C4-C5	27-31	41	C1-C2(C)-C3-C4	196-199	74	C1-C2(C,C)-C(C,C)-C	356-357
9	C1-C2(OH)-C3-C4-C5-C6	32-37	42	C1-C2(C,C)-C	200-201	75	C1-C2-C3-C4-C5-C-C-C-C	358-362
10	C1-C2(C,OH)-C3-C4	38-41	43	C1-C2-C3-C-C-C	202-204	76	C1-C2(C)-C3-C4-C5-C6-C7-C8	363-370
11	C1-C2(C,OH)-C3(C)-C4	42-45	44	C1-C2(C)-C3-C4-C5	205-209	77	C1-C2-C3(C9)-C4-C5-C6-C7-C8	371-379
12	C1-C2(C,OH)-C3(C,C)-C4	46-49	45	C1-C2-C3(C4)-C-C	210-213	78	C1-C2-C3-C4(C9)-C5-C6-C7-C8	380-388
13	HO-C1-C2(C,C)-C3	50-52	46	C1-C2(C,C)-C3-C4	214-217	79	C1-C2(C,C)-C3-C4-C5-C6-C7	389-395
14	C1-C2-C3-C4-C5-C6-C7-OH	53-59	47	C1-C2(C)-C(C)-C	218-219	80	C1-C2(C)-C3(C8)-C4-C5-C6-C7	396-403
15	C1-C2-C3-C4-C5-C6-C7-C8-OH	60-67	48	C1-C2-C3-C4-C-C-C	220-223	81	C1-C2(C)-C3-C4(C8)-C5-C6-C7	404-411
16	C1-C2-C3-C4-C5-C6-C7-C8-C9-OH	68-76	49	C1-C2(C)-C3-C4-C5-C6	224-229	82	C1-C2(C)-C3-C4-C5(C8)-C6-C7	412-419
17	C1-C2-C3-C4-C5-C6-C7-C8-C9-C10-OH	77-86	50	C1-C2-C3(C7)-C4-C5-C6	230-236	83	C1-C2(C)-C3-C4-C-C(C)-C	420-423
18	C1-C2(OH)-C3-C4-C5-C6-C7	87-93	51	C1-C2(C,C)-C3-C4-C5	237-241	84	C1-C2-C3(C8,C)-C4-C5-C6-C7	424-431
19	C1-C2(OH)-C3-C4-C5-C6-C7-C8	94-101	52	C1-C2(C)-C3(C6)-C4-C5	242-247	85	C1-C2-C3(C8)-C4(C9)-C5-C6-C7	432-440
20	C1-C2(OH)-C3-C4-C5-C6-C7-C8-C9-C10	102-111	53	C1-C2(C)-C3-C(C)-C	248-250	86	C1-C2-C3(C5)-C4-C(C)-C-C	441-445
21	C1-C2-C3(OH)-C-C	112-114	54	C1-C2-C3(C4,C)-C-C	251-254	87	C1-C2-C3-C4(C5,C)-C-C-C	446-450
22	C1-C2-C3(OH)-C4-C5-C6	115-120	55	C1-C2-C3(C-C)-C-C	255-257	88	C1-C2-C3(C-C)-C4-C5-C6-C7	451-457
23	C1-C2-C3(OH)-C4-C5-C6-C7	121-127	56	C1-C2(C,C)-C3(C)-C4	258-261	89	C1-C2(C,C)-C3-C4(C7)-C5-C6	458-464
24	C1-C2-C3(OH)-C4-C5-C6-C7-C8	128-135	57	C1-C2-C3-C4-C-C-C-C	262-265	90	C1-C2(C,C)-C3-C4-C5(C)-C6	465-470
25	C1-C2-C3-C4(OH)-C-C-C	136-139	58	C1-C2(C)-C3-C4-C5-C6-C7	266-272	91	C1-C2(C)-C3(C7,C)-C4-C5-C6	471-477
26	C1-C2-C3-C4(OH)-C5-C6-C7-C8	140-147	59	C1-C2-C3(C8)-C4-C5-C6-C7	273-280	92	C1-C2(C)-C3(C7)-C4-C5(C)-C6	478-484
27	C1-C2-C3-C4-C5(OH)-C-C-C-C	148-152	60	C1-C2-C3-C4(C5)-C-C-C	281-285	93	C1-C2(C,C)-C3(C6,C)-C4-C5	485-490
28	C1-C2(C)-C3-OH	153-155	61	C1-C2(C,C)-C3-C4-C5-C6	286-291	94	C1-C2(C,C)-C3(C6)-C4(C)-C5	491-496
29	C1-C2(C,C)-OH	156-157	62	C1-C2(C)-C3(C7)-C4-C5-C6	292-298	95	C1-C2(C,C)-C3-C(C,C)-C	497-499
30	C1-C2-C3(C5)-C4-OH	158-162	63	C1-C2(C)-C3-C4(C7)-C5-C6	299-305	96	C1-C2(C)-C3(C4,C)-C(C)-C	500-503
31	C1-C2(C)-C3-C4-OH	163-166	64	C1-C2(C)-C3-C-C(C)-C	306-308	97	C1-C2(C)-C3(C4-C5)-C(C)-C	504-508
32	C1-C2(C)-C3(OH)-C4	167-170	65	C1-C2-C3(C7,C)-C4-C5-C6	309-315	98	C1-C2-C3(C-C,C-C)-C-C	509-511
33	C1-C2(OH)-C3-C4(C)-C5	171-175	66	C1-C2-C3(C4)-C(C)-C-C	316-319	99	C1-C2-C3-C4-C5-C-C-C-C-C	512-516

Five-element regression equation

Multiple linear regression (MLR) is used to develop four five-element linear models that linked the chemical shift (*CS*) to AEDV and γ parameter for four types of atoms. These models are given by Eq. (3) as stated above. Applying an MLR method and calculating the AEDV and the γ parameter as Eqs. (1) and (2), four models containing five variables are developed for C1, C2, C3 and C4 atom types, respectively. The regression coefficients b_j ($j = 0, 1, 2, 3, 4, 5$) and relative statistic parameters such as correlation coefficient, R,

and root mean squares (*RMS*) error obtained by employment of MLR method are listed in Table 3 (see column M1, M2, M3 and M4). The experimental and estimated chemical shifts *CS* by model M1, M2, M3 and M4 for four atom types are listed in Table 4A, 4B, 4C and 4D, respectively. And the $CS_{estimated}$ plotted vs the $CS_{observed}$ are shown in Fig. 2A, 2B, 2C and 2D.

Table 3 The regression models including 5 descriptors for four types of carbons in alcohols

Statistics	M1	CV1	M2	CV2	M3	CV3	M4	CV4
n	182	181	223	222	78	77	33	32
R	0.9887	0.9887 ± 0.0001	0.9971	0.9971 ± 0.0000	0.9978	0.976 ± 0.0008	0.9968	0.9969 ± 0.0003
RMS	0.906	0.906 ± 0.001	0.821	0.821 ± 0.001	1.091	1.118 ± 0.05	1.009	0.996 ± 0.022
b_0	−4.0842	−4.0932 ± 0.3922	−54.8412	−54.8412 ± 0.0078	−795163	−79.5354 ± 0.0411	−114.8768	−114.8690 ± 0.1785
b_1	47.4927	47.5205 ± 0.3153	35.2941	35.2941 ± 0.0037	34.5150	34.5184 ± 0.0098	35.8260	35.8229 ± 0.0378
b_2	20.5350	20.5436 ± 0.3860	44.2698	44.2698 ± 0.0038	40.3245	40.3224 ± 0.0196	37.2248	37.2243 ± 0.0461
b_3	28.1140	28.1215 ± 0.3806	50.6517	50.6517 ± 0.0039	44.2638	44.2792 ± 0.0178	38.7824	38.8095 ± 0.0810
b_4	34.7681	34.7753 ± 0.3762	54.6307	54.6308 ± 0.0048	46.0407	46.0496 ± 0.0409	41.2137	41.2193 ± 0.0665
γ	−2.9154	−2.9150 ± 0.0015	−2.8354	−2.8354 ± 0.0006	−2.2532	−2.2400 ± 0.0279	0.4925	0.4873 ± 0.0246

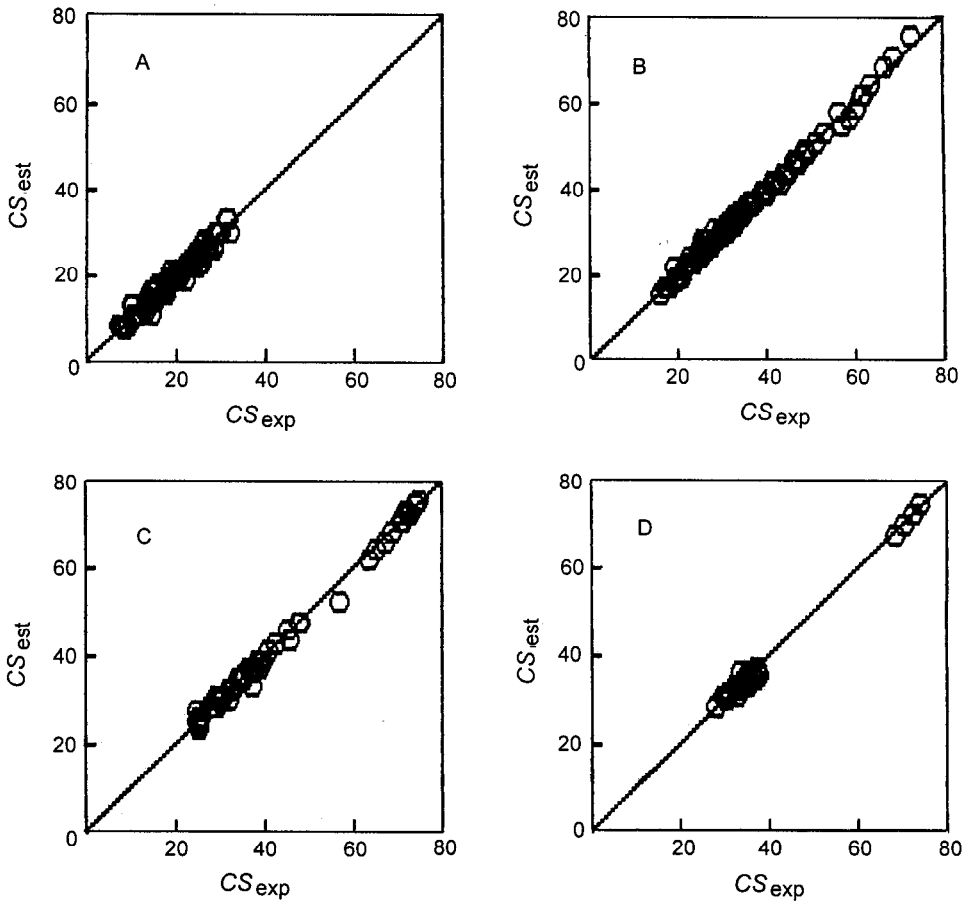

Fig. 2 Plot of the *CS* estimated by M1 (A), M2 (B), M3 (C) and M4 (D) vs the experimental *CS*.

Cross validation

In order to explain the stability of model M1, M2, M3 and M4 and their prediction capacity for chemical shifts of various atoms in external samples, a cross-validation technique, called leave-one-out procedure, is employed to test four models above.

Table 4A *CS* values observed and calculated of various carbon atoms for the 1st type carbon

Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}
1	17.60	17.70	17.70	170	19.70	20.01	20.04	285	19.30	18.92	18.91	404	22.60	22.24	22.23
3	10.00	12.94	12.97	171	24.00	22.74	22.64	286	29.20	29.61	29.63	410	14.00	13.92	13.91
6	13.60	13.90	13.91	175	22.80	22.24	22.23	291	13.90	13.90	13.90	411	19.40	18.94	18.92
10	13.80	13.89	13.89	179	29.80	29.62	29.61	292	18.90	19.48	19.50	412	22.40	22.22	22.21
15	14.20	13.89	13.89	180	17.90	17.26	17.20	297	14.00	13.92	13.92	418	11.00	11.16	11.16
21	25.10	26.01	26.11	183	25.50	25.93	25.95	298	15.10	16.21	16.25	419	19.00	18.95	18.94
23	22.60	22.75	22.76	184	28.50	25.94	25.84	299	22.70	22.24	22.22	420	22.40	22.22	22.21
26	9.90	10.21	10.21	187	15.40	17.19	17.30	304	11.00	11.17	11.17	424	8.00	8.41	8.43
27	23.30	22.72	22.68	189	13.10	13.92	13.93	305	19.00	18.97	18.97	430	13.70	13.90	13.90
31	14.00	13.92	13.92	191	24.30	25.52	25.65	306	22.40	22.22	22.21	431	26.40	26.34	26.34
32	23.30	22.72	22.67	193	13.50	13.89	13.90	309	8.10	8.41	8.42	432	11.90	11.18	11.17
37	13.90	13.90	13.90	196	21.90	22.24	22.25	314	14.80	13.94	13.93	438	14.20	13.92	13.91
38	28.60	30.14	30.30	199	11.50	11.19	11.18	315	26.50	26.34	26.33	439	14.90	16.21	16.25
41	8.50	7.46	7.41	200	31.60	32.91	33.06	316	11.80	11.18	11.17	440	15.30	16.19	16.22
42	26.30	27.41	27.52	202	13.70	13.89	13.90	319	14.80	16.21	16.26	441	11.00	11.17	11.17
45	17.50	15.78	15.68	205	22.70	22.22	22.21	320	10.60	11.13	11.14	445	19.30	18.97	18.95
46	25.40	24.66	24.57	209	14.30	13.92	13.91	325	14.10	13.91	13.91	446	14.90	13.93	13.92
49	25.60	23.18	23.01	210	11.40	11.16	11.16	326	27.10	26.88	26.87	450	27.00	26.32	26.27
52	26.30	28.66	28.75	213	18.70	18.97	18.98	330	13.00	11.21	11.18	451	10.60	11.14	11.14
53	13.90	13.89	13.89	214	28.70	29.64	29.68	331	13.30	13.49	13.50	457	13.70	13.89	13.89
60	13.90	13.89	13.89	217	8.50	8.44	8.44	332	29.90	29.63	29.62	458	29.90	29.64	29.62
68	14.00	13.89	13.89	218	19.20	19.51	19.52	336	24.70	22.26	22.20	463	11.20	11.17	11.17
77	14.00	13.89	13.89	220	13.70	13.89	13.89	337	17.10	16.74	16.72	464	21.90	18.99	18.89
87	22.30	22.71	22.75	224	22.40	22.21	22.21	341	7.90	8.44	8.46	465	29.30	29.61	29.63
93	13.90	13.90	13.90	229	13.60	13.89	13.90	342	23.30	23.63	23.65	470	22.50	22.22	22.21
94	23.40	22.71	22.66	230	10.90	11.16	11.16	343	19.80	19.51	19.50	471	17.10	16.73	16.72
101	14.00	13.89	13.89	235	13.90	13.92	13.92	346	10.40	13.51	13.66	476	14.80	13.94	13.93
102	23.40	22.71	22.66	236	18.80	18.94	18.95	347	19.00	19.46	19.47	477	23.80	23.61	23.60
111	14.00	13.89	13.89	237	29.50	29.61	29.62	351	11.80	11.16	11.15	478	18.90	19.49	19.50
112	9.80	10.18	10.19	241	15.10	13.94	13.92	352	7.50	8.39	8.42	483	22.70	22.24	22.23
115	9.90	10.18	10.18	242	18.90	19.48	19.50	355	23.20	23.10	23.08	484	15.30	16.23	16.26
120	14.00	13.92	13.92	246	11.60	11.19	11.18	356	25.60	24.16	24.08	485	25.60	24.13	24.05
121	10.00	10.18	10.18	247	14.50	16.24	16.29	358	13.80	13.89	13.89	489	9.00	8.46	8.44
127	14.00	13.90	13.90	248	22.70	22.24	22.23	363	22.30	22.21	22.21	490	20.60	20.89	20.91
128	10.00	10.18	10.18	251	7.70	8.42	8.44	370	13.70	13.89	13.89	491	28.20	26.90	26.86
135	13.90	13.89	13.89	254	25.60	26.37	26.41	371	11.10	11.16	11.16	495	20.90	19.53	19.49
136	4.10	13.92	13.92	255	10.50	11.14	11.15	378	13.80	13.89	13.89	496	11.60	10.76	10.68
140	14.00	13.92	13.92	258	27.00	26.91	26.90	379	19.00	18.94	18.94	497	31.80	29.66	29.57
147	14.00	13.90	13.90	261	17.70	16.76	16.72	380	14.00	13.92	13.92	500	17.20	16.76	16.75
148	14.00	13.90	13.90	262	13.60	13.89	13.89	387	13.70	13.90	13.90	503	18.90	20.90	21.04
153	18.90	21.26	21.32	266	22.40	22.21	22.21	388	20.20	18.92	18.86	504	19.50	19.48	19.48
156	31.30	33.41	33.68	272	13.80	13.89	13.89	389	29.20	29.61	29.63	508	14.50	11.18	11.14
158	11.10	11.17	11.17	273	11.30	11.16	11.16	395	13.80	13.90	13.90	509	7.00	8.36	8.41
162	16.00	17.99	18.05	279	14.10	13.90	13.89	396	19.00	19.48	19.49	512	14.00	13.89	13.93
163	22.50	22.22	22.22	280	19.30	18.94	18.93	402	13.80	13.90	13.90				
167	18.10	18.53	18.54	281	14.10	13.92	13.91	403	15.20	16.21	16.24				

In each time, a sample drawn from working data set acts as a external prediction set and the remaining $n-1$ samples construct a calibration set. The calibration set will be used to establish a new model between the *CS* value and the AEDV and parameter(s) γ, and the new model is used to predict the chemical shift of the external set. In the same method, n times of modeling and prediction are executed for working set containing n samples of each atom type. The average results of n regression coefficients, correlation coefficients, and relative statistic parameters in the stage of modeling are listed in Table 3 (see column CV1, CV2, CV3 and CV4). These models are used to predict *CS* of the external set sample. All of prediction results using cross validation technique are listed in Table 4. In Table 4, correlation coefficient (R), root mean squares (*RMS*) error between n predicted and observed *CS* values for four types of atoms are $R=0.9877, 0.9968, 0.9969, 0.9941$ and

RMS =0.945, 0.857, 1.304, 1.375, respectively.

Table 4B *CS* values observed and calculated of various carbon atoms for the 2nd type carbon

Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}
2	57.00	54.73	54.50	92	22.90	22.66	22.65	222	32.00	32.18	32.18	372	29.70	29.59	29.58
4	25.80	25.65	25.64	96	39.60	39.49	39.48	223	29.00	29.51	29.51	374	36.70	36.46	36.46
5	63.60	64.25	64.31	97	26.10	25.88	25.87	226	38.90	39.09	39.09	375	26.90	26.84	26.84
7	19.10	21.66	21.70	98	29.70	29.55	29.55	227	29.70	29.47	29.46	376	32.40	32.20	32.20
8	35.00	35.17	35.18	99	32.20	32.18	32.18	228	23.00	22.67	22.66	377	22.70	22.66	22.66
9	61.40	61.61	61.62	100	22.80	22.66	22.66	231	2.50	29.58	29.58	381	19.40	19.97	19.98
11	22.60	22.65	22.65	104	39.60	39.49	39.49	233	39.00	39.09	39.09	382	9.60	39.11	39.10
12	28.20	31.19	31.23	105	26.20	25.88	25.87	234	20.20	19.96	19.96	384	36.80	36.46	36.45
13	32.50	32.53	32.53	106	30.10	29.58	29.57	239	47.30	46.27	46.21	385	29.30	29.49	29.49
14	61.80	61.63	61.61	107	30.00	29.56	29.55	240	18.10	17.21	17.19	386	23.00	22.67	22.67
16	22.80	22.65	22.65	108	29.60	29.54	29.54	245	26.80	26.87	26.88	391	44.40	43.64	43.60
17	32.00	32.18	32.18	109	32.20	32.18	32.18	250	49.00	48.67	48.65	392	24.40	24.09	24.08
18	25.80	28.55	28.58	110	22.90	22.65	22.65	252	33.40	34.09	34.12	393	33.00	32.21	32.20
19	32.80	32.55	32.55	113	29.70	29.93	29.94	256	25.20	26.91	26.96	394	22.80	22.66	22.66
20	61.90	61.63	61.61	116	30.30	29.96	29.95	263	22.70	22.65	22.65	399	34.00	33.75	33.74
25	32.00	32.58	32.60	118	39.40	39.46	39.46	264	32.10	32.18	32.18	400	30.00	29.51	29.50
29	41.60	42.11	42.13	119	19.40	19.00	18.99	265	29.40	29.53	29.53	401	23.10	22.67	22.67
30	19.10	18.98	18.97	122	29.70	29.96	29.96	268	39.30	39.11	39.11	406	47.00	46.04	45.99
34	39.20	39.46	39.46	124	36.90	36.82	36.82	269	27.20	26.82	26.81	408	39.90	39.13	39.11
35	28.30	28.50	28.51	125	28.20	28.53	28.53	270	32.40	32.19	32.19	409	19.90	19.97	19.97
36	22.90	22.67	22.67	126	23.00	22.68	22.67	271	22.80	22.65	22.65	414	36.50	36.42	36.42
40	36.50	37.11	37.15	129	30.30	29.96	29.95	274	29.70	29.58	29.58	415	34.40	33.75	33.73
50	72.60	75.72	76.11	131	37.20	36.83	36.83	276	36.50	36.44	36.44	417	29.50	29.59	29.59
54	22.80	22.65	22.65	132	25.70	28.72	28.76	277	29.70	29.49	29.48	422	39.50	39.13	39.12
55	32.10	32.18	32.18	133	32.30	32.21	32.20	278	23.30	22.67	22.66	423	25.20	24.14	24.09
56	29.40	29.53	29.53	134	22.90	22.66	22.65	282	20.20	19.96	19.96	425	34.20	34.11	34.11
57	26.10	28.57	28.60	137	19.10	19.00	19.00	283	39.50	39.11	39.10	427	41.30	40.97	40.96
58	32.90	32.55	32.55	138	40.00	39.48	39.47	288	44.10	43.62	43.59	428	26.40	26.76	26.77
59	61.90	61.63	61.61	141	19.10	19.01	19.00	289	27.00	26.74	26.73	429	23.70	22.69	22.67
61	22.80	22.65	22.65	142	40.30	39.49	39.47	290	23.70	22.68	22.66	433	26.70	26.90	26.90
62	32.10	32.18	32.18	144	37.50	36.83	36.82	295	36.70	36.40	36.39	436	36.50	36.42	36.42
63	29.60	29.54	29.54	145	28.20	28.53	28.53	296	0.70	19.98	19.97	437	20.80	19.98	19.97
64	29.70	29.55	29.55	146	23.00	22.68	22.67	301	46.60	46.02	45.99	442	30.00	29.61	29.60
65	26.10	28.57	28.61	149	23.00	22.68	22.68	303	29.90	29.60	29.59	444	44.40	43.38	43.32
66	32.90	32.55	32.55	150	28.30	28.53	28.53	308	36.90	36.40	36.39	447	17.30	17.24	17.24
67	61.90	61.63	61.61	151	37.50	36.84	36.83	310	34.30	34.11	34.10	448	44.80	43.64	43.59
69	22.90	22.65	22.65	155	68.90	71.18	71.40	312	44.30	43.62	43.58	452	25.60	26.94	26.97
70	32.20	32.18	32.18	159	25.90	28.60	28.65	313	17.30	17.24	17.23	454	32.70	33.80	33.83
71	29.60	29.53	29.53	161	66.90	68.54	68.69	317	26.70	26.89	26.90	455	29.20	29.51	29.51
72	29.90	29.55	29.55	165	41.80	42.11	42.12	321	25.60	26.93	26.97	456	23.10	22.67	22.67
73	29.80	29.56	29.55	166	60.20	58.92	58.81	323	35.40	36.44	36.46	460	51.00	50.55	50.53
74	26.20	28.57	28.60	173	48.90	49.04	49.05	324	20.00	19.98	19.98	462	31.00	29.62	29.59
75	32.90	32.55	32.55	176	58.90	56.19	55.90	329	24.40	24.15	24.14	467	42.00	40.93	40.88
76	62.00	61.63	61.60	177	46.40	46.64	46.65	334	53.30	53.20	53.19	468	33.90	33.68	33.67
78	22.80	22.65	22.65	188	15.90	15.75	15.74	340	32.60	31.40	31.34	474	43.10	40.93	40.82
79	32.20	32.18	32.18	190	24.90	25.27	25.28	350	22.60	24.22	24.29	475	17.00	17.25	17.26
80	29.60	29.54	29.54	194	22.20	2.63	22.63	353	30.60	31.45	31.50	481	43.90	43.33	43.30
81	29.90	29.56	29.55	195	34.10	34.80	34.83	359	22.70	22.65	22.65	488	28.80	28.68	28.67
82	29.80	29.56	29.56	198	31.60	32.21	32.23	360	32.00	32.18	32.18	499	56.50	57.73	57.98
83	29.80	29.56	29.56	203	22.70	22.65	22.65	361	29.40	29.53	29.54	507	21.10	21.54	21.58
84	26.10	28.57	28.61	204	31.70	32.15	32.16	362	29.60	29.55	29.55	510	27.10	28.80	28.93
85	32.90	32.55	32.55	207	41.90	41.74	41.73	365	39.20	39.11	39.11	513	22.80	22.65	22.65
86	61.90	61.63	61.61	208	20.80	19.94	19.92	366	27.40	26.84	26.83	514	32.30	32.18	32.18
89	39.50	39.48	39.48	211	29.40	29.56	29.56	367	29.70	29.55	29.55	515	29.80	29.53	29.53
90	25.80	25.86	25.86	216	36.50	36.74	36.75	368	32.00	32.18	32.19	516	30.10	29.55	29.55
91	32.30	32.20	32.19	221	22.60	22.65	22.65	369	22.70	22.65	22.65				

Table 4C *CS* values observed and calculated of various carbon atoms for the 3rd type carbon

Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}
22	63.40	62.15	61.94	172	65.20	64.23	64.15	294	38.50	39.09	39.11	413	28.40	28.30	28.29
24	68.70	68.50	68.48	174	24.80	25.36	25.39	300	25.40	26.13	26.16	416	34.80	34.64	34.64
28	67.00	66.38	66.34	181	74.80	75.29	75.36	302	32.10	32.47	32.48	421	28.10	28.28	28.29
33	67.20	66.40	66.34	186	84.70	88.43	90.84	307	28.40	28.29	28.29	434	39.40	39.11	39.09
44	38.80	37.54	37.39	192	25.00	24.03	23.88	318	39.00	39.09	39.09	435	36.80	36.98	36.99
88	67.20	66.40	66.35	197	29.90	30.38	30.41	322	40.60	40.96	41.00	443	32.00	2.48	32.50
95	67.20	66.40	66.35	206	27.90	28.26	28.28	328	45.40	43.52	43.29	453	40.70	40.98	41.01
103	67.20	66.40	66.35	212	36.80	36.73	36.72	335	25.30	23.94	23.82	461	31.90	30.29	30.13
114	73.80	74.84	74.94	219	34.00	34.86	34.92	338	35.10	35.06	35.05	469	28.90	28.30	28.28
117	72.30	72.73	72.76	225	28.10	28.28	28.28	344	29.80	30.60	30.65	472	35.00	35.07	35.08
123	72.60	72.75	72.76	232	34.30	34.61	34.62	345	45.30	45.68	45.76	479	32.40	32.77	32.79
130	72.60	72.75	72.76	243	31.90	32.74	32.78	348	29.10	30.63	30.73	480	36.20	36.94	36.99
139	70.60	70.62	70.62	244	40.60	41.21	41.24	349	47.60	47.55	47.55	482	25.70	26.15	26.17
143	70.90	70.63	70.60	249	25.70	26.12	26.13	364	28.00	28.28	28.29	493	47.90	48.00	48.01
152	71.10	70.65	70.60	257	42.40	43.08	43.18	373	34.60	34.63	34.63	494	27.40	28.42	28.56
154	30.80	30.74	30.74	260	37.90	37.17	37.09	383	32.60	32.51	32.51	501	37.10	32.91	32.29
160	37.50	37.09	37.06	267	28.10	28.28	28.29	397	32.20	32.76	32.79	505	29.00	28.48	28.41
164	24.80	27.50	27.59	275	34.70	34.63	34.63	398	38.80	39.11	39.12	506	56.80	52.03	51.66
168	35.10	35.22	35.23	284	32.30	32.50	32.51	405	25.30	26.14	26.17				
169	72.00	72.97	73.11	293	32.80	32.76	32.76	407	30.20	30.35	30.36				

Table 4D *CS* values observed and calculated of various carbon atoms for the 4th type carbon

Na	CS_{exp}	$CS_{M}O$	CS_{CV}	Na	CS_{exp}	CS_{MO}	$CSCV$	Na	CS_{exp}	CS_{MO}	CS_{CV}	Na	CS_{exp}	CS_{MO}	CS_{CV}
39	70.30	69.95	69.81	201	28.00	28.43	28.58	339	34.90	34.46	34.40	486	36.00	36.06	36.07
43	72.20	72.07	72.01	215	30.30	30.38	30.40	354	34.80	34.30	34.17	487	37.30	37.45	37.50
47	74.10	75.06	75.71	238	30.60	30.95	30.97	357	35.00	35.50	35.64	492	34.00	33.63	33.57
48	37.50	35.88	35.40	253	32.30	32.35	32.35	390	30.20	30.96	31.01	498	32.40	32.09	32.00
51	32.60	30.76	30.52	259	32.70	32.51	32.47	426	32.50	32.92	32.95	502	33.60	36.58	38.44
157	68.40	67.99	67.80	287	30.10	30.96	31.02	449	32.80	33.47	33.58	511	37.10	36.26	35.91
178	29.70	31.15	31.25	311	32.80	32.91	32.92	459	31.00	31.53	31.58				
182	35.00	32.88	32.54	327	33.00	33.07	33.07	466	30.10	30.97	31.03				
185	36.90	34.59	33.61	333	30.90	31.51	31.58	473	35.50	35.02	34.96				

It should be indicated that it is essential to do cross validation in the development of a new quantitative structure-spectrum relationship (QSSR) model. We added two compounds, CH_3OH (CS_{exp} = 49.00) and CH_3CH_3 (CS_{exp} = 5.70), into the model M1 and established the relationship between the *CS* and the AEDV and the γ parameter. The results showed that model M1 had a good estimation capacity being R = 0.9898 and *RMS* = 0.927 but had a poor prediction capacity. It was found that the square of correlation coefficient was negative value and the prediction values of CH_3OH and CH_3CH_3 were 83.23 and 40.33 after cross validation had been done. When the compound CH_3OH was deleted from working data set, and R was 0.9884 and *RMS* was 0.926 but the *CS* value of CH_3CH_3 predicted by cross-validation was 43.41. When the compound CH_3CH_3 was deleted from working data set, and R was 0.9901 and *RMS* was 0.910 but the *CS* of CH_3OH predicted by cross-validation was 95.86. A main reason of poor results for CH_3OH and CH_3CH_3 is that their atoms belong to type 1 and AEDV of these atoms only have one element (a_1) whereas AEDV of atoms belonging to type 1 for the other compounds have more elements than one.

γ-*Effect*

The γ-effect worldwide exists in the ^{13}C NMR spectrum. If the γ-effect is not considered in modeling

linear models and four-parameter models are directly developed by only using AEDV, the errors estimated and the errors predicted of models are larger than those of five-parameter models. The correlation coefficients for four types of carbons are $R = 0.9227$, 0.9734, 0.9916 and 0.9963, respectively. And relative *RMS* are 2.33, 2.48, 2.16 and 1.10, respectively. Aparting from the fourth type of carbon, there are significant differences between four-parameter and five-parameter models. This is because the γ-effect of the fourth type of carbon is little than that of the other three types of carbons.

Chemical shifts for unknown compounds

Six alcohols and alkanes, HO-C1-C-OH (No.1), HO-C1-C2-C-OH (No.2), C1-C2(OH)-C3(C6)-C4-C5 (No.3), C1-C2-C3(C5)-C4-C-C(C)-C-C (No.4), C1-C2-C3-C4-C5(C9,C)-C6-C7-C8 (No.5) and C1-C2-C3-C4-C5-C6-C-C-C-C-C (No.6), are taken to be an external prediction set. The 29 experimental values of chemical shifts of various carbon atoms to be predicted are taken from Ref. 24. These compounds are not included in previous modeling data set. So, the prediction ability of the five-parameter models for unknown compounds can be demonstrated well. The experimental CS_{exp} and the CS_{cal} predicted by models M1, M2, M3 and M4 for these six compounds in external prediction set are listed in Table 5.

Table 5 *CS* values of experiment and predicted by five-parameter models for 6 unknown compounds

Na	No.1		No.2		No.3		No.4		No.5		No.6	
	CS_{exp}	CS_{prd}	CS_{exp}	CS_{prd}	CS_{exp}	CS_{prd}	CS_{exp}	CS_{prd}	CS_{exp}	CS_{prd}	CS_{exp}	CS_{prd}
1	63.4	64.6	59.0	60.6	18.5	20.0	11.4	11.2	14.2	13.9	14.2	13.9
2			35.2	35.5	69.3	70.9	29.7	29.6	23.9	22.7	23.0	22.7
3					40.6	41.6	34.3	34.6	26.6	26.8	32.3	32.2
4					24.4	25.9	35.1	33.8	42.1	41.0	29.7	29.5
5					10.8	11.2	19.3	18.9	32.9	33.5	30.1	29.6
6					12.9	15.3			45.0	43.6	30.1	29.6
7									17.5	17.2		
8									15.1	13.9		
9									27.5	26.3		

Conclusion

Four models which link the AEDV vector and γ parameter to *CS* for all four type carbon atoms have been developed by using 63 alkanes and 36 alcohols as the calibration set. To demonstrate the prediction ability of these models for unknown external prediction set, the *CS* values of various chemically nonequivalent carbon atoms in six external molecules have been calculated and the cross validation technique has been performed with very good satisfactory results. It is foreseeable that the AEDV vector and its correlation models would be utilized widely in quantitative structure-spectra relationship (QSSR) studies. The researches are in progress.

References

1. Jurs, P.C.; Ball, J.W.; Anker, L.S.; Friedman, T. L., *J. Chem. Inf. Comput. Sci.*, **32**, 272(1992).
2. Miyashita, Y.; Okuyama, T.; Ohsako, H.; Sasaki, S., *J. Am. Chem. Soc.*, **110**, 3469 (1989).
3. Miyashita, Y.; Li, Z.; Sasaki, S., *Trend. Anal. Chem.*, **12**(2), 50 (1993).
4. Wiener, H., *J. Am. Chem. Soc.*, **69**, 17 (1947).
5. Hosoya, H., *Bull. Chem. Soc. Jap.*, **44**, 2332 (1971).
6. Randic, M., *J. Am. Chem. Soc.*, **97**, 6609 (1975).
7. Bolaban, A.T., *Pure Appl. Chem.*, **55**, 199 (1983).
8. Schultz, H.P., *J. Chem. Inf. Comput. Sci.*, **29**, 227 (1989).
9. Kier, L.B.; Hall, L.H.; Murray, W., *J. Pharm. Sci.*, **64**, 1971 (1975).
10. Randic, M., *J. Math. Chem.*, **7**, 155 (1991).
11. Randic, M., *J. Chem. Inf. Comput. Sci.*, **26**, 134 (1986).
12. Grant, D.M.; Paul, E.G., *J. Am. Chem. Soc.*, **86**, 2984(1964).
13. Lindeman, L.P.; Adams, J.Q., *Anal. Chem.*, **43**,

1245(1971).
14. Small, G. W.; Jurs, P. C., *Anal. Chem.*, **56**, 1313 (1984).
15. Small, G. W.,; Stouch, T. R.; Jurs, P. C., *Anal. Chem.*, **56**, 2341 (1984).
16. Kvasnicka, V., *J. Math. Chem.*, **6**, 63 (1991).
17. Doucet, J. P.; Panaye, A.; Feuilleaubois, E.; Ladd, P., *J. Chem. Inf. Comput. Sci.*, **33**, 320 (1993).
18. Anker, L. S.; Jurs, P. C., *Anal. Chem.*, **64**, 1157 (1992).
19. Svozil, D.; Pospichal, J.; Kvasnicka, V., *J. Chem. Inf. Comput. Sci.*, **35**, 924 (1995).
20. Clouser, D. L.; Jurs, P. C., *Anal. Chim. Acta*, **321**, 127 (1995).
21. Liu, S. S.; Cao, C. Z.; Li, Z., *Chin. J. Chem.*, **16**, 126(1998).
22. Aylward, G. H.; Findlay, T. J. V., *SI Chemical Data*, John Wiley & Sons, 2nd Ed, 1982.
23. Roberts, J. D.; Weigert, F. J.; Kroschwitz, J. I.; Reich, H. J., *J. Am. Chem. Soc.*, **92**, 1338(1970).
24. Sadtler research laboratories, division of bio-rad laboratories, Inc. researchers, *Sadtler Standard Carbon*-13 NMR *Spectra*, Editor & publishers, 1979-1980.

(E9906065 JIANG, X. H.; DONG, L. J.)