Journal of Chromatography, 630 (1993) 223-230
Elsevier Science Publishers B.V., Amsterdam

CHROM. 24 712

Structure-gas chromatographic retention time models of tetra-n-alkylsilanes and tetra-n-alkylgermanes using topological indexes

Eugene J. Kupchik

Department of Chemistry, St. John's University, Grand Central and Utopia Parkways, Jamaica, NY 11439 (USA)

(First received August 4th, 1992; revised manuscript received October 27th, 1992)

ABSTRACT

Structure-gas chromatographic retention time models were developed for 26 tetra-n-alkylsilanes, 26 tetra-n-alkylgermanes, and the mixed set of silanes and germanes using topological indexes. The topological indexes used were molecular connectivity indexes, the new Kier-Hall total topological indexes, and the new Kier-Hall electrotopological state atom index, S, for the silicon and germanium atoms. For the mixed data set, a model based on ${}^1\chi$, ${}^3\chi_c$ and S gave r = 0.999 and s = 0.033. In two cases the residual was more than twice the standard error. Replacement of ${}^1\chi$ by TTV, the new Kier-Hall total topological index using δ^v values for the silicon and germanium atoms, also gave $r = 0.999$ and s = 0.033 but in no case was there a residual greater than twice the standard error. A two-variable model based on ${}^0\chi^r$, which uses δ^r values for the metal atoms, and ${}^4\chi_{pc}{}^v$, which uses δ^v values for the metal atoms gave r = 0.999 and s = 0.035. The results demonstrate the adequacy of the Kier-Hall electrotopological state atom index for encoding the atomic characteristics of silicon and germanium in a mixed data set. The results show that the Kier-Hall total topological indexes can be used in place of ${}^1\chi$ to encode the size and skeletal branching of silanes and germanes.

INTRODUCTION

Topological indexes are numerical descriptors of molecules that are based on certain topological features of their hydrogen-suppressed graphs [1–6]. Topological indexes are used to correlate the structures of molecules with their physical, chemical, or biological properties. Probably the most widely used topological indexes are the molecular connectivity indexes. Molecular connectivity was extensively developed by Kier and Hall [7–11] from the alkane branching index of Randić [12]. The method describes the structure of a molecule by a set of molecular connectivity indexes, χ, which are calculated from a hydrogen-suppressed structural formula or graph of the molecule. The simplest index, ${}^0\chi$, is given by:

$$ {}^0\chi = \Sigma\, \delta_i^{-1/2} \qquad (1)$$

where the sum is over all non-hydrogen bonded atoms in the molecule and δ_i is a number assigned to each atom, which equals the number of atoms connected to it. The first order molecular connectivity index, ${}^1\chi$, which is often called the Randić branching index, is given by:

$$ {}^1\chi = \Sigma(\delta_i\delta_j)^{-1/2} \qquad (2)$$

where the sum is over all of the edges of the graph whose two atoms have the δ values shown. The second order molecular connectivity index, ${}^2\chi$, is given by:

$$ {}^2\chi = \Sigma(\delta_i\delta_j\delta_k)^{-1/2} \qquad (3)$$

where the sum is over all paths of length two whose three atoms have the δ values shown. Higher order

---

Correspondence to: E. J. Kupchik, Department of Chemistry, St. John's University, Grand Central and Utopia Parkways, Jamaica, NY 11439, USA.

0021-9673/93/$06.00

path connectivity indexes, ${}^{m}\chi_{p}$, where p denotes path and m denotes path length, can also be calculated. In addition to paths, a molecule can be dissected into cluster and path-cluster fragments. ${}^{3}\chi_{c}$ is a third order cluster fragment given by:

$$ {}^{3}\chi_{c} = \Sigma(\delta_i\delta_j\delta_k\delta_l)^{-1/2} \tag{4} $$

where c denotes cluster and the sum is over all isobutane fragments whose four atoms have the δ values shown. ${}^{4}\chi_{pc}$ is a fourth order path-cluster index given by:

$$ {}^{4}\chi_{pc} = \Sigma\,(\delta_i\delta_j\delta_k\delta_l\delta_m)^{-1/2} \tag{5} $$

where pc denotes path-cluster and the sum is over all isopentane fragments whose five atoms have the δ values shown.

For non-sp^3 carbon atoms and heteroatoms, Kier and Hall [13–16] have suggested the use of δ^{v} (valence δ) values, instead of δ values. The use of δ^{v} gives a valence molecular connectivity index, ${}^{m}\chi_{t}^{v}$, where m is the order of the index and t is the type of index, namely path, cluster or path-cluster. δ^{v} is obtained from the following equation:

$$ \delta^{v} = \frac{Z^{v} - h}{Z - Z^{v} - 1} \tag{6} $$

where Z^{v} is the number of valence electrons of the atom, h is the number of attached hydrogen atoms, and Z is the atomic number of the atom. For an sp^3 carbon, $\delta^{v} = \delta$. As an alternate approach for encoding information about the heteroatom in heteroalkanes, Kupchik [17–19] has suggested that the heteroatom be assigned a radius $\delta\,(\delta^{r})$ value obtained from the following equation:

$$ \delta^{r} = \frac{r_c}{r_{het}}\,\delta \tag{7} $$

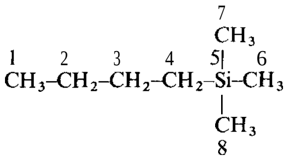
where r_c is the covalent radius of carbon, r_{het} is the covalent radius of the heteroatom and δ is the number of sp^3 carbon atoms attached to the heteroatom. The use of eqn. 7 gives a radius molecular connectivity index, ${}^{m}\chi_{t}^{r}$. Because it is based on experimentally determined covalent radii, δ^{r} is empirical and probably less attractive than δ^{v} which is essentially non-empirical. In this paper both δ^{v} and δ^{r} are employed and the results compared.

Recently Hall and Kier [20] introduced a new highly discriminating index called the total topological index. This index is obtained by adding the entries that appear in their novel topological state matrix of a molecule. Each entry in the matrix can be calculated using the formula

$$ t_{ij} = \frac{GM_{ij}}{d_{ij}^{2}} \tag{8} $$

TABLE I

TOPOLOGICAL STATE MATRIX OF BUTYLTRIMETHYLSILANE (δ^{v}_{Si} = 0.44)



$$ \overset{1}{CH_3}-\overset{2}{CH_2}-\overset{3}{CH_2}-\overset{4}{CH_2}-\overset{5}{Si}(\overset{7}{CH_3})(\underset{8}{CH_3})-\overset{6}{CH_3} $$

	1	2	3	4	5	6	7	8	Topological state index T_i
1	1.000	0.354	0.176	0.105	0.051	0.034	0.034	0.034	1.788
2		2.000	0.500	0.222	0.086	0.051	0.051	0.051	3.315
3			2.000	0.500	0.134	0.072	0.072	0.072	3.526
4				2.000	0.235	0.106	0.106	0.106	3.380
5					0.440	0.166	0.166	0.166	1.444
6						1.000	0.085	0.085	1.599
7							1.000	0.085	1.599
8								1.000	1.599

Total topological index (TTV) = 14.345

where GM_{ij} is the geometric mean of the δ values of the atoms in the path of atoms of length d_{ij} atoms between atoms i and j; t_{ij} gives the topological information for the vertex sequence beginning with atom i and ending with vertexj. Alternatively, other algorithms based on GM_{ij} and d_{ij} may be used to give alternate values of t_{ij}. The topological state matrix of butyltrimethylsilane is given in Table I. All entries were calculated using a hand-held calculator. To illustrate the use of eqn. 8 the value of 0.222 for $t_{2,4}$ is calculated as follows:

$$t_{2\,4} = \frac{(2 \cdot 2 \cdot 2)^{1/3}}{3^2}$$

The sum of all of the entries in the matrix gives the total topological index value of 14.345. In this paper this index is given the symbol TTV, the V indicating that the δ values used were obtained using eqn. 6; the index is denoted by TTS if only simple δ values were used; the index is denoted by TTR if the δ values used were obtained from eqn. 7. Although total topological indexes can be obtained using a hand-held calculator, they are more easily obtained using the MOLCONN-X computer program[a]. The topological state index T_i for atom i is obtained by adding the entries in column i and row i of the topological state matrix. Atoms having the same topological state index are topologically equivalent.

Hall *et al.* [21,22] recently introduced a new atom index called the electrotopological state index. In this method each atom in the hydrogen-suppressed graph is assigned an intrinsic state value, I_i, obtained from the following equation:

$$I_i = [(2/N)^2 \delta^{\mathrm{v}} + 1]/\delta \qquad (9)$$

where N is the principal quantum number of atom i, δ^{v} is the valence δ value of atom i, and δ is the simple δ value of atom i. The electrotopological state index for an atom i, S_i, is defined as follows:

$$S_i = I_i + \Delta I_i \qquad (10)$$

where ΔI_i is a perturbation term arising from the electronic interaction between each atom in the molecule. The perturbations are summed over the entire molecule:

$$\Delta I_i = \frac{\Sigma\,(I_i - I_j)}{r_{ij}^2} \qquad (11)$$

where r_{ij} is the graph separation between atoms i and j, counted as numbers of atoms, including i and j. Electrotopological state calculations for butyltrimethylsilane using a hand-held calculator are given in Table II. The electrotopological index value of $-$ 0.6792 for the silicon atom is obtained by adding the Z value for silicon (0.694) to the AZ value for silicon ($-$ 1.3732). Electrotopological state values for atoms are easily obtained with the MOLCONN-X program.

Only a few studies employing organometallic compounds and topological indexes have been reported [1618,231. Further studies of organometallic compounds and other organoelement compounds using topological indexes are needed to test the adequacy of the valence δ formalism and to test the new total topological$ndexes and the new electrotopological state atom index. The purpose of the present study was to determine if satisfactory structure-gas chromatographic retention time models could be found for some organosilicon and organogermanium compounds using molecular connectivity indexes, total topological indexes and electrotopological state indexes.

MATERIALS AND METHODS

Meaningful structure-property relationship studies require high-quality property data. This study uses the high-quality gas chromatographic retention time data of Semlyen *et al.* [24]. These workers obtained gas chromatographic retention time data for 26 tetra-n-alkylsilanes and 26 tetra-*n*-alkylgermanes. Analysis on columns of 2–13% squalene on Embacel at 100°C gave essentially similar results. Retention data were expressed as logarithms of retention times relative to mesitylene = 100, and agreement to within ± 0.01 was obtained between chromatograms.

All of the topological indexes used in this study were obtained using the MOLCONN-X program. The statistical analyses were performed using the SYSTAT statistical program [25]. The multiple regression equations presented in this paper appear to be free of the problem of multicollinearity, or near-linear dependence among the regression vari-

---

[a] From L. H. Hall, Hall Associates Consulting, 2 Davis Street, Quincy, MA 02170, USA.

ables [26,27]. Regression models fit to data by the method of least squares when strong multicollinearity is present are very poor predictor equations. Intercorrelation data is presented with each multiple regression equation to demonstrate a non-multicollinearity problem. Other diagnostic tests employed to demonstrate non-multicollinearity were an inspection of the condition indices of the correlation matrix and an inspection of the tolerance of each variable. A condition index greater than 15 indicates a possible problem and one greater than 30 suggests a serious problem with multicollinearity [25,27]. Values of tolerance near zero indicate that some of the predictors are highly intercorrelated [25]. The multiple regression equations presented in this paper passed these tests.

RESULTS AND DISCUSSION

The 26 silanes in Table III were used in the stepwise regression procedure of the SYSTAT statistical program [25]. The following topological indexes were included as final candidates in the regression equation: $^0\chi, ^1\chi, ^2\chi, ^3\chi, ^4\chi, ^3\chi_c, ^4\chi_c, ^4\chi_{pc}$, the corresponding set of valence indexes, TTS, TTV, TTR and S for the silicon atom for a total of 20 indexes. No compound had a zero value for any index. Higher order path indexes were not used to avoid zero values for compounds. The following one-variable equation was obtained:

$$\text{Log } T = -0.426(0.055) + 0.507(0.010)^1\chi \qquad (12)$$

$$n = 26 \qquad r = 0.996 \qquad s = 0.062 \qquad F = 2822$$

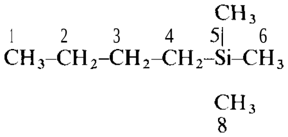
T is the retention time; the numbers in parentheses are the standard errors of the coefficients; n is the number of compounds, r is the correlation coefficient, s is the standard error of estimate, and F is the variance ratio. In only one case was there a residual greater than twice the standard error. Also the following two-variable equation was obtained:

$$\text{Log } T = -0.067(0.059) + 0.483(0.006)^1\chi - 0.251(0.035)^3\chi_c \qquad (13)$$

$$n = 26 \qquad r = 0.999 \qquad s = 0.035 \qquad F = 4441$$

TABLE II

ELECTROTOPOLOGICAL STATE CALCULATIONS FOR BUTYLTRIMETHYLSILANE

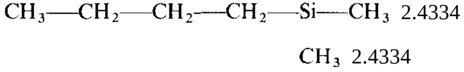


$$\overset{1}{CH_3}-\overset{2}{CH_2}-\overset{3}{CH_2}-\overset{4}{CH_2}-\underset{\underset{8}{CH_3}}{\overset{\overset{CH_3}{|}}{\overset{5}{Si}}}-\overset{6}{CH_3}$$

Intrinsic values: $I(1) = 2.000$, $I(2) = 1.500$, $I(3) = 1.500$, $I(4) = 1.500$, $I(5) = 0.694$, $I(6) = 2.000$, $I(7) = 2.000$, $I(8) = 2.000$.

i	$(I_i - I_j)/r_{ij}^2$ Matrix								
	1	2	3	4	5	6	7	8	ΔI_i = row sum
1	0.0	0.1250	0.0556	0.0313	0.0522	0.0000	0.0000	0.0000	0.2641
2	-0.1250	0.0	0.0000	0.0000	0.0504	-0.0200	-0.0200	-0.0200	-0.1346
3	-0.0556	0.0000	0.0	0.0000	0.0896	-0.0313	-0.0313	-0.0313	-0.0599
4	-0.0313	0.0000	0.0000	0.0	0.2015	-0.0556	-0.0556	-0.0556	0.0034
5	-0.0522	-0.0504	-0.0896	-0.2015	0.0	-0.3265	-0.3265	-0.3265	-1.3732
6	0.0000	0.0200	0.0313	0.0556	0.3265	0.0	0.0000	0.0000	0.4334
7	0.0000	0.0200	0.0313	0.0556	0.3265	0.0000	0.0	0.0000	0.4334
8	0.0000	0.0200	0.0313	0.0556	0.3265	0.0000	0.0000	0.0	0.4334
									0.0000

$S_i = I_i + \Delta I_i$



$$\overset{2.2641}{CH_3}-\overset{1.3654}{CH_2}-\overset{1.4401}{CH_2}-\overset{1.5034}{CH_2}-\underset{CH_3\ 2.4334}{\overset{\overset{CH_3\ 2.4334}{|}}{Si}}\overset{-0.6792}{}-CH_3\ 2.4334$$

TABLE III

LOGARITHMS OF RETENTION TIMES OF TETRA-n-ALKYLSILANES AND TETRA-n-ALKYLGERMANES AND TOPOLOGICAL INDEXES

No.	Compound[a]	$^1\chi$	$^3\chi_c$	S	TTV	Log T[b]	Calc.[c]	Res.[d]
1	Me_3BuSi	3.5607	1.5607	-0.6782	14.3584	1.29	1.29	0.00
2	$MeEt_3Si$	3.6820	0.9286	-0.6713	14.1734	1.43	1.49	-0.06
3	Me_2Pr_2Si	4.1213	1.2071	-0.7033	17.2755	1.60	1.64	-0.04
4	$Me_2EtBuSi$	4.1213	1.2071	-0.6982	17.2958	1.61	1.64	-0.03
5	$MeEt_2PrSi$	4.1820	0.9268	-0.6973	17.2185	1.70	1.73	-0.03
6	Et_4Si	4.2426	0.7071	-0.6914	17.1717	1.80	1.80	0.00
7	$Me_2PrBuSi$	4.6213	1.2071	-0.7243	20.3554	1.90	1.88	0.02
8	$MeEtPr_2Si$	4.6820	0.9268	-0.7234	20.2999	1.95	1.97	-0.02
9	Et_3PrSi	4.1426	0.7071	-0.7174	20.2666	2.05	2.05	0.00
10	Me_2Bu_2Si	5.1213	1.2071	-0.7453	23.4573	2.15	2.12	0.03
11	$MePr_3Si$	5.1820	0.9268	-0.7494	23.4178	2.20	2.21	-0.01
12	Et_2Pr_2Si	5.2426	0.7071	-0.7434	23.3979	2.29	2.29	0.00
13	Et_3BuSi	5.2426	0.7071	-0.7384	23.3913	2.32	2.29	0.03
14	$MePr_2BuSi$	5.6820	0.9268	-0.7704	26.5620	2.47	2.45	0.02
15	$MeEtBu_2Si$	5.9067	1.1309	-0.7098	27.9648	2.49	2.49	0.00
16	$EtPr_3Si$	5.7426	0.7071	-0.7695	26.5656	2.52	2.53	-0.01
17	Pr_4Si	6.2426	0.7071	-0.7955	29.7698	2.74	2.71	-0.03
18	$EtPr_2BuSi$	6.2426	0.7071	-0.7905	29.7462	2.77	2.77	0.00
19	Et_2Bu_2Si	6.2426	0.7071	-0.7854	29.7254	2.82	2.77	0.05
20	$MeBu_3Si$	6.6820	0.9268	-0.8123	32.9167	2.99	2.93	0.06
21	Pr_3BuSi	7.0985	1.1154	-0.7609	33.9562	2.99	3.07	-0.08
22	$EtPrBu_2Si$	6.7426	0.7071	-0.8114	32.9489	3.03	3.01	0.02
23	Pr_2Bu_2Si	7.2426	0.7071	-0.8375	36.2088	3.24	3.25	-0.01
24	$EtBu_3Si$	7.2426	0.707 1	-0.8324	36.1137	3.30	3.25	0.05
25	$PrBu_3Si$	7.1426	0.7071	-0.8584	39.4616	3.49	3.49	0.00
26	Bu_4Si	8.2426	0.7071	-0.8794	42.7365	3.72	3.13	-0.01
27	Me_3BuGe	3.5607	1.5607	-1.1086	13.4109	1.42	1.41	0.01
28	$MeEt_3Ge$	3.6820	0.9268	-1.1250	13.0229	1.57	1.61	-0.04
29	Me_2Pr_2Ge	4.1213	1.2071	-1.1597	16.1384	1.77	1.77	0.00
30	$Me_2EtBuGe$	4.1213	1.2071	-1.1503	16.1933	1.77	1.76	0.01
31	$MeEt_2PrGe$	4.1820	0.9286	-1.1632	15.9902	1.84	1.86	-0.02
32	Et_4Ge	4.2426	0.7071	-1.1667	15.8485	1.94	1.94	0.00
33	$Me_2PrBuGe$	4.62 13	1.2071	-1.1885	19.1749	2.04	2.01	0.03
34	$MeEtPr_2Ge$	4.6820	0.9268	-1.2014	18.9886	2.09	2.10	-0.01
35	Et_3PrGe	4.7426	0.7071	-1.2049	18.8569	2.18	2.18	0.00
36	Me_2Bu_2Ge	5.1213	1.207 1	-1.2172	22.2309	2.31	2.25	0.06
37	$MePr_3Ge$	5.1820	0.9268	-1.2396	22.0178	2.35	2.35	0.00
38	Et_2Pr_2Ge	5.2426	0.7071	-1.2431	21.8963	2.42	2.43	-0.01
39	Et_3BuGe	5.2426	0.7071	-1.2336	21.9310	2.45	2.42	0.03
40	$MePr_2BuGe$	5.6820	0.9268	-1.2683	25.1095	2.61	2.59	0.02
41	$MeEtBu_2Ge$	5.9067	1.1309	-1.2033	26.4804	2.64	2.63	0.01
42	$EtPr_3Ge$	5.7426	0.7071	-1.2813	24.9666	2.66	2.67	-0.01
43	Pr_4Ge	6.2426	0.7071	-1.3194	28.0678	2.89	2.92	-0.03
44	$EtPr_2BuGe$	6.2426	0.7071	-1.3100	28.0891	2.91	2.91	0.00
45	Et_2Bu_2Ge	6.2426	0.7071	-1.3006	28.1125	2.95	2.91	0.04
46	$MeBu_3Ge$	6.6820	0.9268	-1.3258	31.3514	3.14	3.07	0.07
47	Pr_3BuGe	7.0985	1.1154	-1.2926	32.1469	3.13	3.22	-0.09
48	$EtPrBu_2Ge$	6.7426	0.7071	-1.3388	31.2313	3.16	3.16	0.00
49	Pr_2Bu_2Ge	1.2426	0.7071	-1.3769	34.3810	3.38	3.40	-0.02
50	$EtBu_3Ge$	7.2426	0.7071	-1.3675	34.3930	3.40	3.40	0.00
51	$PrBu_3Ge$	1.7426	0.7071	-1.4057	37.5670	3.61	3.64	-0.03
52	Bu_4Ge	8.2426	0.7071	-1.4344	40.7725	3.85	3.89	-0.04

[a] Me = Methyl, Et = ethyl, Pr = n-propyl, Bu = n-butyl.

[b] Logarithm of retention time taken from ref. 24.

[c] Calculated from eqn. 20.

[d] Residual = Log T- calc.

The intercorrelation of $^1\chi$ and $^3\chi_c$ in this equation is: r = 0.521, which is quite small, indicating that these variables are generally independent. In only one case was there a residual greater than twice the standard error. Eqn. 13 is a significant improvement over eqn. 12. The $^1\chi$ index in eqns. 12 and 13 encodes the size of the molecule and the skeletal branching. In the case of each silane molecule there are four four-atom fragments contributing to the $^3\chi_c$ index, each of which includes the silicon atom. The improvement obtained by adding $^3\chi_c$ to eqn. 12 demonstrates that the environment around the silicon atom is important in determining the retention time. Substitution of either TTV or TTR for $^1\chi$ in eqn. 13 gave improved results in that in no case was there a residual greater than twice the standard error. With TTV, $r = 0.999$ and $s = 0.032$, while with TTR, $r = 0.999$ and $s = 0.035$. TTV, which uses a δ value for silicon obtained from eqn. 6, and TTR, which uses a δ value for silicon obtained from eqn. 7, encode the silicon atom as well as the size and skeletal branching in the molecule. Substitution of TTS, which uses a simple δ value for silicon, for $^1\chi$ in eqn. 13 did not improve eqn. 13 ($r = 0.997$, $s = 0.050$).

When only the valence connectivity indexes were allowed to be candidates in the regression equation, stepwise regression gave the following two-variable equation:

$$\text{Log } T = -0.979(0.056) + 0.349(0.004)^0\chi^v - 0.391(0.040)^4\chi_c^v \qquad (14)$$

$n = 26$ $r = 0.999$ $s = 0.029$ $F = 6610$

In no case was there a residual greater than twice the standard error. The intercorrelation between $^0\chi^v$ and $^4\chi_c^v$ is: $r = 0.594$. When only the radius connectivity indexes were allowed to be candidates in the regression equation, a two-variable equation of equal quality was obtained:

$$\text{Log } T = -0.670(0.053) + 0.349(0.004)^0\chi^r - 0.954(0.098)^4\chi_c^r \qquad (15)$$

$n = 26$ $r = 0.999$ $s = 0.029$ $F = 6610$

In no case was there a residual greater than twice the standard error. The intercorrelation between $^0\chi^r$ and $^4\chi_c^r$ is: $r = 0.594$. In eqns. 14 and 15, $^0\chi^v$ and $^0\chi^r$ mostly encode atom identity and are only moderately sensitive to skeletal branching; $^4\chi_c^v$ and $^4\chi_c^r$, which consist of one five-atom fragment containing the silicon atom, encode both atom identity and the environment around the silicon atom. These results further demonstrate that the structural and electronic environment around the silicon atom is important in determining the retention time.

The 26 germanes in Table III were also used in the stepwise regression procedure of the SYSTAT program. The same 20 topological indexes were used as in the case of the silanes. No compound had a zero value for any index. The following two-variable equation was obtained:

$$\text{Log } T = 0.067(0.057) + 0.481(0.006)^1\chi - 0.230(0.034)^3\chi_c \qquad (16)$$

$n = 26$ $r = 0.999$ $s = 0.034$ $F = 4633$

In two cases the residual was twice the standard error. The intercorrelation between $^1\chi$ and $^3\chi_c$ is: $r = 0.522$. Substitution of TTV or TTR for $^1\chi$ in eqn. 16 gave equations of higher quality. With TTV ($r = 0.999$, $s = 0.034$) in no case was there a residual greater than twice the standard error, while with TTR ($r = 0.999$, $s = 0.034$) in only one case was there a residual greater than twice the standard error.

As in the case of the silanes, when only the valence connectivity indexes were allowed to be candidates in the regression equation, an excellent two-variable equation was obtained:

$$\text{Log } T = -1.210(0.055) + 0.348(0.004)^0\chi^v - 0.214(0.021)^4\chi_c^v \qquad (17)$$

$n = 26$ $r = 0.999$ $s = 0.026$ $F = 7772$

In no case was there a residual greater than twice the standard error. The intercorrelation between $^0\chi^v$ and $^4\chi_c^v$ is: $r = 0.594$. An equation of equal quality was obtained when only the radius connectivity indexes were allowed to be candidates in the regression equation:

$$\text{Log } T = -0.529(0.049) + 0.347(0.004)^0\chi^r - 0.880(0.089)^4\chi_c^r \qquad (18)$$

$n = 26$ $r = 0.999$ $s = 0.027$ $F = 7615$

In no case was there a residual greater than twice the standard error. The intercorrelation between $^0\chi^r$ and $^4\chi_c^r$ is: $r = 0.594$. Eqns. 17 and 18 are analogous to eqns. 14 and 15 obtained for the silanes, which is not surprising, since the silanes and germanes differ only in the nature of the metal atom.

It was of special interest to study a mixed set of the 26 silanes and 26 germanes. A mixed data set always provides a more stringent test for the adequacy of topological indexes. As in the cases of the silanes and germanes, the mixed set of 26 silanes and 26 germanes in Table III were used in the stepwise regression procedure of the SYSTAT program. The same topological indexes were included as final candidates in the regression equation, namely $^0\chi$, $^1\chi$, $^2\chi$, $^3\chi$, $^4\chi$, $^3\chi_c$, $^4\chi_c$, $^4\chi_{pc}$, the corresponding set of valence indexes, TTS, TTV, TTR and S for the metal atom for a total of 20 indexes. The following two-variable equation was obtained:

$$\text{Log } T = -0.434(0.036) + 0.478(0.007)^1\chi + 0.022(0.002)^3\chi^v \quad (19)$$

$n = 52 \quad r = 0.997 \quad s = 0.052 \quad F = 3521$

In two cases the residual was more than twice the standard error. The intercorrelation between $^1\chi$ and $^3\chi^v$ is: $r = -0.429$. The addition of $^0\chi$ to eqn. 19 lowered s to 0.035 but intercorrelation between $^1\chi$ and $^0\chi$ was unsatisfactory *($r = -0.996$)*.

It was desirable to obtain an equation involving the electrotopological atom index, S, to see if it could adequately encode the silicon and germanium atoms in the mixed data set. Accordingly, only the simple connectivity indexes, namely $^0\chi$, $^1\chi$, $^2\chi$, $^3\chi$, $^4\chi$, $^3\chi_c$, $^4\chi_c$, and $^4\chi_{pc}$, and the electrotopological state atom index, S, for the metal atom were allowed to be final candidates in the regression equation. The following excellent three-variable equation was obtained:

$$\text{Log } T = -0.215(0.041) + 0.468(0.004)^1\chi - 0.228(0.023)^3\chi_c - 0.281(0.018)S \quad (20)$$

$n = 52 \quad r = 0.999 \quad s = 0.033 \quad F = 6749$

The results of this analysis are given in Table III. In two cases the residual was more than twice the standard error. Multicollinearity is not a problem as is evidenced by the following intercorrelation results: $r = 0.212$ for S and $^1\chi$, $r = -0.034$ for S and $^3\chi_c$, and $r = 0.502$ for $^1\chi$ and $^3\chi_c$. In eqn. 20, S encodes the metal atom, $^1\chi$ encodes size and skeletal branching, and $^3\chi_c$ encodes the structural environment around the silicon atom. The replacement of $^1\chi$ in eqn. 20 by TTV ($r = 0.999$, $s = 0.033$) or TTR ($r = 0.999$, $s = 0.035$) gave improved results in that in no case was there a residual greater than twice the standard error. The replacement of $^1\chi$ in eqn. 20 by TTS gave a poorer result ($s = 0.046$).

Finally, the following excellent two-variable equation was obtained when only the valence connectivity indexes and the radius connectivity indexes were allowed to be candidates in the regression equation:

$$\text{Log } T = -1.255(0.031) + 0.371(0.003)^0\chi^r + 0.039(0.002)^4\chi^v_{pc} \quad (21)$$

$n = 52 \quad r = 0.999 \quad s = 0.035 \quad F = 8820$

In only one case was there a residual greater than twice the standard error. The intercorrelation between $^0\chi^r$ and $^4\chi^v_{pc}$ is: $r = -0.008$. Eqn. 21 is excellent because the two connectivity indexes are essentially orthogonal, which is an ideal result in a multivariate analysis. Th problem of the non-orthogonality of descriptors has been addressed by Randić [28-31]. In eqn. 21, $^0\chi^r$ encodes mostly atom identity; $^4\chi^v_{pc}$ has several five-atom fragments contributing to it each of which contains the metal atom; it encodes the structural and electronic environment around the metal atom, further demonstrating that this environment is important in determining the retention time.

The results reported in this paper further demonstrate that molecular connectivity indexes can be successfully used to develop structure-property models for organosilicon and organogermanium compounds. The results especially demonstrate the adequacy of the new Kier-Hall electrotopological state atom index for encoding the atomic characteristics of silicon and germanium in a mixed data set. It has also been shown that the new Kier-Hall highly discriminating total topological index can be used in place of the less discriminating $^1\chi$ index to encode both size and skeletal branching in molecules. The adequacy of the δ^v values and δ^r values for the silicon and germanium atoms has also been demonstrated.

REFERENCES

1 D. H. Rouvray, ***Sci. Am., 255 (1986) 40.***

2 D. Bonchev, ***Information Theoretic Indices for Characterization of Chemical Structures,*** Wiley, New York, 1983.

3 N. Trinajstić, *Chemical Graph Theory, Vol. 11,* CRC Press, Boca Raton, FL, 1983, pp. 105–140.

4 P. J. Hansen and P. C. Jurs, *J. Chem. Educ., 65 (1988) 574.*

5 P. G. Seybold, M. May and U. A. Bagal, *J. Chem. Educ., 64 (1987) 575.*

6 A. Sabljić, in W. Karcher and J. Devillers (Editors), *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology,* ECSC, EEC, EAEC, Brussels and Luxembourg, 1990, pp. 61–82.

7 L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research,* Academic Press, New York, 1976.

8 L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis,* Wiley, New York, 1986.

9 L. B. Kier, in S. H. Yalkowsky, A. A. Sinkula and S. C. Valvani (Editors), *Physical Chemical Properties of Drugs,* Marcel Dekker, New York, 1980, Ch. 9.

10 L. B. Kier, W. J. Murray, M. Randic and L. H. Hall, *J. Pharm. Sci., 65 (1976) 1226.*

11 L. B. Kier and L. H. Hall, *Quant. Struct.–Act. Relat.,* 1 0 (1991) 134.

12 M. RandiC, *J. Am.* Chem. *Soc.*, 97 (1975) 6609.

13 L. B. Kier and L. H. Hall, *J. Pharm. Sci., 70 (1981) 583.*

14 L. B. Kier and L. H. Hall, *J. Pharm. Sci., 72 (1983)* 1170.

15 L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis,* Wiley, New York, 1986, pp. 15–20.

16 L. H. Hall and D. Aaserud, *Quant. Struct.–Act. Relat., 8 (1989) 296.*

17 E. J. Kupchik, *Quant. Struct.–Act. Relat., 5 (1986) 95.*

18 E. J. Kupchik, *Quanf. Struct.–Act. Relat., 7 (1988) 57.*

19 E. J. Kupchik, *Quant. Struct.–Act. Relat., 8 (1989) 98.*

20 L. H. Hall and L. B. Kier, *Quant. Struct.–Act. Relat.*, 9 (1990) 115.

21 L. H. Hall, B. Mohney and L. B. Kier, *Quant. Struct.–Act. Relat.*, 10 (1991) 43.

22 L. H. Hall, B. Mohney and L. B. Kier, *J. Chem. Inf. Comput. Sci.*, 31 (1991) 76.

23 E. J. Kupchik, *Quanf. Struct.–Act. Relat., 4 (1985) 123.*

24 J. A. Semlyen, G. R. Walker, R. E. Blofeld and C. S. G. Phillips, *J. Chem. Soc., (1964) 4948.*

25 L. Wilkinson, *SYSTAT: The System for Statistics,* SYSTAT, Evanston, IL, 1988.

26 D. C. Montgomery and E. A. Peck, *Introduction to Linear Regression Analysis,* Wiley, New York, 2nd ed., 1992.

27 S. Chatterjee and B. Price, *Regression Analysis by Example*, Wiley, New York, 2nd ed., 1991.

28 M. Randić, *J. Chem. Inf. Comput. Sci.,* 31 (1991) 311.

29 M. RandiC, New *J.* Chem., 15 (1991) 517.

30 M. RandiC, *J. Comput. Chem., 12 (1991) 970.*

31 M. Randic, *J. Chem. Educ.*, 69 (1992) 713.