

A Novel Quantum-Topology Index

Xinhua Li,* Qingsen Yu, and Longguan Zhu

Unit Research into Molecular Connectivity and Drug Design, Department of Chemistry, Zhejiang University, Hangzhou, Zhejiang 310027, P. R. China

Received June 22, 1999

By modifying the δ values of molecular connectivity indexes, new quantum-topology descriptors are presented in this paper. Their ability to describe molecular structure, both geometric and electronic, is established by correlating them with the bonding energy (E), force constant (κ), and radius (R) of a heterogeneous set of monohydrides, as well as with the solubility in water $-\log S$, hydrophobic constants $-\log P$, and molecular volumes of halogenobenzenes and the standard heat of formation of SiX_n ($n = 1-4$, $\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{I}$). Moreover, it is clearly demonstrated that this ability is greater than that shown by the χ connectivity indexes.

INTRODUCTION

Molecular topological indexes and quantum chemistry parameters, two important descriptors in studying structure–activity relationships, are used to characterize the constitution and configuration of molecules. Molecular topological indexes are obtained from molecular ichnographic–topological structure, which is relatively impersonal and simple.¹ Molecular topological indexes are calculated based on only the framework of a molecule, but some nonbonding forces are important in QSAR or QSPR. Quantum-chemistry parameters are obtained from molecular configuration in quantum level, which can reveal all the electronic information of the molecule in principle, for example, electron distribution probability, energy, polarity, and electronic charge density, etc. But experimental hypotheses were often used in practical calculations, and some information on molecular structure would then be lost and result in some limitations. However, the vital short coming of the quantum chemistry method is that it is inconvenient to deal with the geometric effect, such as the geometric characters and bulk of molecules, while geometric effect is very important in QSAR. Then Bader and his colleagues presented a theory of quantum-topology in 1977,² which may unify the molecular orbital theory and the topological theory. Meanwhile Jiang Yuansheng and Xu Lu also presented their own quantum-topology indexes.³

Molecular connectivity indexes are a kind of widely used topological index and have been proved to be successful in many fields, but the limitations of molecular connectivity index are obvious. Their physical meanings are ambiguous, and they cannot distinguish stereoisomers. There also have been various ways try to modify the molecular connectivity index.^{4–8} However, comparing the two methods above, we find that they are complementary. If the two methods above are combined, they can offset some disadvantages, respectively. So we connect the molecular connectivity and quantum chemistry together by modifying the base of molecular connectivity— δ value.

METHODOLOGY

The delta value (δ) was defined by Randic in his branching indexes; it was used to express the number of bonds connected with a certain carbon atom. Randic considered that when the δ value of a bonding carbon atom in an alkyl isomer was different, the corresponding bond was different, too. The concept has been further developed and applied extensively by Randic, Kier, Hall, and their co-workers; they modified the effect of the nonsaturated bond, heteroatoms, and ringed structure to the Randic branching indexes and presented the molecular connectivity indexes.^{1–4} But there are still some disadvantages of the δ value in reflecting electric structure of heteroatoms. Comparing the molecular connectivity (adjacent) matrix with the Mulliken population matrix, we find that they have almost the same form, and the δ values of a hydrocarbon have a good relationship with its Mulliken overlap populations. For example, five hydrocarbons' Mulliken overlap populations were calculated by using 6-31G** basis sets at the Hartree–Fock level. The sum of Mulliken overlap populations of each carbon and its δ value have a correlation constant of 0.999; the calculation results and correlated equation are shown in Table 1 and eq 1, where, δ_i is the δ value of each carbon atom, and δ_i^* is the sum of Mulliken overlap populations of each carbon atom, s is the standard deviation of errors, f is the F statistic for correlation, and r is the correlation coefficient.

Table 1. δ Values and Mulliken Overlap Populations of Some Hydrocarbons

		C1	C2	C3	C4	C5
ethane	δ_i	1.0000	1.0000			
	δ_i^*	0.3311	0.3311			
propane	δ_i	1.0000	2.0000	1.0000		
	δ_i^*	0.3429	0.6858	0.3429		
butane	δ_i	1.0000	2.0000	2.0000	1.0000	
	δ_i^*	0.3443	0.6961	0.6961	0.3443	
isopropane	δ_i	1.0000	3.0000	1.0000	1.0000	
	δ_i^*	0.3551	1.0653	0.3551	0.3551	
neopropane	δ_i	1.0000	4.0000	1.0000	1.0000	1.0000
	δ_i^*	0.3692	1.47692	0.3692	0.3692	0.3692

* To whom all correspondence should be addressed.

$$\delta_i = 0.0843 + 2.6927\delta_i^* \quad (1)$$

$$s = 0.05, \quad f = 4366.6, \quad r^2 = 0.998$$

Galvez also found that the distribution of charges between atoms has a good relationship with molecular properties.⁸ The δ value describes the number of chemical bonds in a molecule, so the δ value describes information on bonding. The Mulliken overlap population can reflect information on bonding, and then weighting the δ value by Mulliken overlap population is reasonable. Heteroatoms have great polarity, but the overlap electrons were distributed to atoms on average by Mulliken in the Mulliken population, so we cannot weigh heteroatoms according to that of carbon; furthermore, the heteroatoms' δ values also have some disadvantages. Aimed at addressing the disadvantages of heteroatoms' δ values, we have introduced a formula to modify the δ values, which has been verified as being reasonable;⁹ that is

$$\delta_i^v = \frac{Z_i^v(\sigma_i - h_i)}{N^2} \quad (2)$$

where δ_i^v is the δ value of the i th atom, z_i^v is the number of valence electrons outside the nucleus, h_i is the number of hydrogens forming σ bonds with the atom, σ_i is the number of bonding electrons of the atom, and N is the highest quantum number of the atom. The definition of δ value can be seen as the ability of the nuclear effect on the extranuclear bonding electrons (exclude those electrons forming σ bonds with hydrogen), so we can replace the $(\sigma_i - h_i)$ term by the Mulliken overlap population. According to eq 2, the δ values of carbon are consistent with that of molecular connectivity index, so the advantages of molecular connectivity in dealing with hydrocarbon can be maintained. Then we can weigh the δ values by Mulliken overlap population as below

$$\delta_A^v = \frac{Z_A}{N_A^2} (2 \sum_{\lambda}^B \sum_{\mu}^A \sum_i n_i c_{A\mu i}^* c_{B\lambda i} s_{A\mu, B\lambda}) \quad (3)$$

$$\delta_B^v = \frac{Z_B}{N_B^2} (2 \sum_{\lambda}^B \sum_{\mu}^A \sum_i n_i c_{A\mu i}^* c_{B\lambda i} s_{A\mu, B\lambda}) \quad (4)$$

where $2 \sum_{\lambda}^B \sum_{\mu}^A \sum_i n_i c_{A\mu i}^* c_{B\lambda i} s_{A\mu, B\lambda}$ are the Mulliken overlap populations between atom A and atom B, δ_A^v and δ_B^v are the valence values of δ of atom A and B, and Z_A and Z_B are available valence charges of the nucleus of atom A and B; they can be approximately defined as

$$Z_A = Z^n - Z^m \quad (5)$$

where Z^n and Z^m terms are the charges of the nucleus and the electrons of the atomic kernel of atom A respectively. Equation 3 can be seen as the contribution of atom A to bonding, while eq 4 is the contribution of atom B; their sum ($N(A, B)$) is more correlated with the strength of the A–B bond than the Mulliken overlap population ($n(A, B)$). For example, 11 monohydrides were calculated by using 6-31G** basis sets at the Hartree–Fock level, and electronic correlation effects are also considered through MP2/6-31G**. The results are shown in Table 2, and for comparison the

Table 2. Values of E ($\times 4.2$ J/mol), κ (N/cm), and R (Å)

no.	molecule	$n(A, B)$		$N(A, B)$		E	κ	R
		HF	RMP2	HF	RMP2			
1	HF	0.2869	0.2784	0.7891	0.7657	135.0	8.80	0.92
2	HCl	0.2661	0.2659	0.4731	0.4728	103.0	4.80	1.27
3	HBr	0.3097	0.3091	0.4452	0.4443	87.4	3.80	1.41
4	HO	0.2747	0.2700	0.6868	0.6751	111.0	7.79	0.97
5	HS	0.2546	0.5481	0.4243	0.4247	82.3	4.20	1.34
6	HC	0.2417	0.2414	0.4833	0.4829	99.0	4.37	1.12
7	HSi	0.2617	0.2617	0.3781	0.3780	71.0	2.48	1.52
8	HB	0.2393	0.2393	0.4187	0.4189	79.0	3.03	1.23
9	HAL	0.2953	0.2954	0.3938	0.3938	68.0	1.62	1.65
10	HN	0.2616	0.2593	0.5887	0.5835	93.0	6.00	1.04
11	HP	0.2535	0.2537	0.3944	0.3947	82.0	3.30	1.42

Table 3. Result of Regression Analysis of Table 2

calculation level	$y = kx + b$		slope, k	intercept, b	correlation coefficient r
	y	x			
HF/6-31G**	$n(A, B)$	E	135.33	55.65	0.15
	$N(A, B)$	E	132.63	25.86	0.91
	$n(A, B)$	κ	13.37	0.98	0.13
	$N(A, B)$	κ	15.82	−3.31	0.97
	$n(A, B)$	R	1.93	0.75	0.18
	$N(A, B)$	R	−1.52	2.02	0.87
RMP2/6-31G**	$n(A, B)$	E	43.42	80.32	0.05
	$N(A, B)$	E	140.05	22.69	0.92
	$n(A, B)$	κ	2.40	3.92	0.02
	$N(A, B)$	κ	16.73	−3.70	0.97
	$n(A, B)$	R	3.13	0.43	0.29
	$N(A, B)$	R	−1.61	2.06	0.88

experimental values of bonding energy (E), force constant (κ), and radius (R) are also listed. In the table, $n(A, B)$ is the Mulliken overlap population, and $N(A, B)$ is the sum of values calculated by eqs 3 and 4. The correlation analysis of corresponding parameters is shown in Table 3. From Tables 2 and 3 we can see that Mulliken overlap population cannot be correlated with any parameter well, but our modified δ value can be correlated well with all three parameters, especially with the force constant. It means that our modified formulas 3 and 4 are more reasonable than Mulliken overlap population as an indication of bond strength.

According to eqs 3 and 4, as N and Z_A of carbon are 2 and 4, the modified δ value of carbon is just the Mulliken overlap population; this is consistent with the fact that the Mulliken overlap population of carbon has a good relationship with its δ value. Then it is clear that the modified δ values reflect the information on bonds between atoms.

Through the molecular connectivity (adjacent) matrix and Mulliken population matrix, we also can see that there are many population matrix parallelisms with one molecular connectivity (adjacent) matrix, and they are just topological equivalents. Atomic state and chemical bond can be affected by the chemical environment enclosing the atom; when the chemical environments changed the atomic state and chemical bond changed, too. Although the atomic state and chemical bond change in value, they are equivalent in topology if they are topologically equivalent. It consistent with topological theory, but molecular connectivity index cannot reflect this property, while quantum chemistry can reflect it, and banding them together can compensate for some disadvantages of them and also convert molecular connectivity index into a novel quantum-topology index.

The modified molecular connectivity indexes are calculated still by the original definition. That is

$${}^mX_t = \sum_{j=1}^n \prod_{i=1}^{m+1} (\delta_i)_j^{-1/2} \quad (6)$$

where n is the number of subfigures, m is the rank of subfigures, δ is the atomic delta value, and t is the type of subfigure.

METHODS OF CALCULATION

All the molecules were optimized at the 6-31G**, STO-3G, and AM1 levels in restricted Hartree-Fock formalism (RHF) field (while the effects of flexible bond to the original conformation of molecules were considered sufficiently). The electronic populations between atoms and molecular volumes were calculated by the Gaussian 94 program, and molecular connectivity indexes were calculated by our own program.¹⁰ All the calculations were carried out on a compatible Pentium 686 microcomputer.

RESULTS AND DISCUSSION

Hydrophobic constant $-\log P$ is an important parameter in pharmacology and environment field,¹¹ but it is difficult to determine by experiment. To compare our quantum-topology indexes with molecular connectivity indexes, 21 molecules¹² with many kinds of polar groups were calculated by RHF/6-31G**. The correlated results of our quantum-topology indexes and molecular connectivity indexes with the hydrophobic constant $-\log P$ show that our quantum-topology indexes are obviously improved compared to molecular connectivity indexes. This means that the weighted δ values are more reasonable than that of the original δ values in describing the electronic structure of polar molecules, and they offset the disadvantages of molecular topological indexes. The correlation equations of molecular connectivity indexes and our modified indexes with the hydrophobic constant are shown in eqs 7 and 7*, and the calculation results are listed in Table 4, where n is the number of molecules.

$$\log P = 0.5040 - 1.2013^0 \chi_P + 2.0123^1 \chi_P + 0.9578^2 \chi_P \quad (7)$$

$$n = 21, \quad S = 0.47, \quad f = 19.4, \quad r^2 = 0.774$$

$$\log P = -1.4730 + 0.3476^1 X_P \quad (7^*)$$

$$n = 21, \quad S = 0.37, \quad f = 102.5, \quad r^2 = 0.844$$

It is well-known that the calculation results of ab initio methods are precise, but they too require time. As to molecules in the same system, the precision of results is not necessary. So long as the results can reflect the relative trend of diversification, they can be accepted. Then if calculations in low-level or semiempirical methods are adopted, much time can be saved. The standard heat of formations of 16 SiX_n ($n = 1-4$, X = F, Cl, Br, I) in ref 11 were calculated by RHF/STQ-3G, while solubility in water $-\log S$ and hydrophobic constants $-\log P$ of 36 halogenobenzenes¹¹

were also calculated by semiempirical method AM1 at the same time. The result show that our modified indexes have good correlation with all these properties; this means that calculations in low-level or semiempirical methods are available. The correlation equations are shown in eqs 8–10.

$$\Delta H_f^0 = 428.4825 - 565.2827^0 X_P + 164.8343 \Delta^0 X_P \quad (8)$$

$$n = 16, \quad s = 144.66, \quad f = 82.6, \quad r^2 = 0.927$$

$$-\log S = -0.7647 + 1.7138^0 X_P + -1.3962^1 X_P \quad (9)$$

$$n = 36, \quad s = 0.40, \quad f = 233.6, \quad r^2 = 0.934$$

$$-\log P = 0.6932 + 1.3079^0 X_P + -1.0796^1 X_P \quad (10)$$

$$n = 36, \quad s = 0.15, \quad f = 898.6, \quad r^2 = 0.982$$

The vital inconvenience of the quantum chemistry method is that it cannot explain the geometric effect, but geometric effect is often an important factor in QSAR, while topological indexes have advantages in dealing with the geometric effect. Molecular volumes of the two systems above are also calculated by Gaussian 94, and from comparison with molecular connectivity indexes we can see that our modified indexes keep the advantage of molecular connectivity in dealing with geometric effect and offset the disadvantage of quantum chemistry index at the same time. The correlation equations of molecular connectivity indexes and our modified indexes with the molecular volumes are shown in eqs 11, 11*, 12, and 12*.

$$V_m = 22.0027 + 5.6267^1 \chi_P \quad (11)$$

$$n = 16, \quad s = 5.11, \quad f = 335.4, \quad r^2 = 0.960$$

$$V_m = -19.7689 + 7.4012^0 X_P \quad (11^*)$$

$$n = 16, \quad s = 5.93, \quad f = 244.8, \quad r^2 = 0.946$$

$$V_m = 62.2494 + 58.2688^3 \chi_P \quad (12)$$

$$n = 36, \quad s = 6.10, \quad f = 212.2, \quad r^2 = 0.852$$

$$V_m = 33.9330 + 4.3810^0 X_P \quad (12^*)$$

$$n = 36, \quad s = 5.84, \quad f = 234.9, \quad r^2 = 0.874$$

Banding molecular connectivity theory and molecular orbital theory together offsets not only the disadvantage of molecular connectivity indexes in dealing with steric conformation but also some novel theories of quantum chemistry, such as solvent effect, can be used in quantum-topology index. This provides a molecular model almost as close to practical molecular structure as QSAR, and makes the result of molecular design more reasonable.

CONCLUSIONS

Through modifying the δ value of molecular connectivity index, we band molecular orbital theory and topological theory together effectively and convert molecular connectiv-

Table 4. Calculation Results of Eqs 7 and 7*

molecule	log P^a	error ^b	log P^c	error	log P_{exptl}^d
methylamine	-0.23	-0.34	-0.30	-0.27	-0.57
acetic acid	0.04	-0.21	-0.16	-0.01	-0.17
formic acid	-0.27	-0.27	-0.50	-0.04	-0.54
acetone	0.30	-0.54	-0.07	-0.17	-0.24
propylaldehyde	0.54	0.05	0.24	0.35	0.59
diethyl ether	0.67	0.22	1.18	-0.29	0.89
1-propyl alcohol	0.82	-0.48	0.62	-0.28	0.34
2-propyl alcohol	0.76	-0.71	0.36	-0.31	0.05
allyl alcohol	0.30	-0.13	0.26	-0.09	0.17
ethanol	0.28	-0.59	0.13	-0.44	-0.31
dimethyl ether	-0.36	0.46	0.30	-0.20	0.10
methanol	-0.33	-0.33	-0.17	-0.49	-0.66
chloroform	1.80	0.17	2.12	-0.150	1.97
tetrachloromethane	2.70	0.13	2.77	0.06	2.83
difluoromethane	-0.08	0.28	0.00	0.20	0.20
dichloromethane	1.02	0.23	1.44	-0.19	1.25
fluoromethane	-0.39	0.90	-0.26	0.77	0.51
chloromethane	0.22	0.69	0.61	0.30	0.91
1,1-difluoroethane	0.24	0.51	0.32	0.43	0.75
acetonitrile	-0.16	-0.18	-0.34	0.00	-0.34
1,1-dichloroethane	1.64	0.15	0.97	0.82	1.79

^a Calculated by eq 7. ^b Error, absolute error. ^c Calculated by eq 7*. ^d log P_{exptl} , observed data.

ity index into a novel quantum-topology index. From practical calculations we can see that the novel quantum-topology indexes are obviously improved in dealing with molecular electronic structure as compared to original indexes, and they have both advantages of topological indexes and quantum parameters and offset some disadvantages of topological indexes and quantum parameters.

ACKNOWLEDGMENT

We thank Dr. Li-Hwa Lu and Dr. Shang Zhi Cai for computers and programs.

REFERENCES AND NOTES

- (1) Wang, L. S.; Zhi, Z. L. *Molecular Connectivity and Molecular Structure-Activity*; Environmental Scientific Press of China: Beijing, 1992; p 12.
- (2) Bader, R. F. W. *Quantum Topology. Proc. Int. Congr. Quantum Chem.* **1980**, 3, 57-61.
- (3) Xu, L. *Chemometrical Method*; Scientific Press: Beijing, 1985; p 304.
- (4) Hall, L. H.; Kier, L. B. Determination of Topological Equivalence in Molecular Graphs from the Topological State. *Quant. Struct-Act. Relat.* **1990**, 9, 115-122.
- (5) Pogliani, L. Molecular Connectivity Model for Determination of Isoelectric Point of Amino Acids. *J. Pharm. Sci.* **1992**, 81 (4), 334-336.
- (6) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Wiley, Research Studies Press, Ltd.: Letchworth, England, 1986; p 32.
- (7) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976; p 21.
- (8) Galvez, J.; Garcia, R.; Salabert, M. T.; Soler, R. Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 520-525.
- (9) Li, Xinhua; Zhu, Longguan; Yu, Qingsen, A Novel Delta Value for Hetero-atoms in Molecular Connectivity. *Acta Chim. Sin.* (Chinese), in press.
- (10) Zhu, L. G.; Ph.D. Thesis; Zhejiang University, Hangzhou, 1995; p 13.
- (11) Yiang, F.; Yan, X. C.; Ou, Y. L.; et al. The improved valence connectivity topological index and its application δ_i^- comparison with δ^v in QSAR/QSPR of three systems. *Chin. J. Chem. Phys.* **1998**, 11, 221-225.
- (12) Sangster, J. Octanol-Water Partition Coefficients of Simple Organic Compounds. *J. Phys. Chem. Ref. Data* **1989**, 18 (3), 1111-1215.

CI9900581