Chem. Pharm. Bull.
36(12)4821—4824(1988)

Comparison of Novel Descriptors σ_{S° and μ^2/α with Topological Indices in Quantitative Structure–Activity Relationships Analyses

HIDEKO KAWAKI*,[a] and YOSHIO SASAKI[b]

Faculty of Pharmacy, Kinki University,[a] Kowakae 3-4-1, Higashi-Osaka 577, Japan and Faculty of Pharmaceutical Sciences, Osaka University,[b] Yamadaoka 1-6, Suita 565, Japan

(Received April 14, 1988)

The shape index $^n\kappa$ ($n=1, 2, 3$) proposed by Kier, as one of the topological indices, as well as the branching index (B.I.) and molecular connectivity χ, have been used as effective descriptors in quantitative structure–activity relationships (QSAR) analyses. These three descriptors are non-energetic and belong to the category of the entropy of information, but not that of thermodynamics. They are linearly related with S°_{298}(g), rather than σ_{s°, indicating the contributions of both dispersion and repulsive forces.

The results of regression analyses of the binding of substituted phenyl glycosides to concanavalin A using the novel QSAR descriptors σ_{s° and μ^2/α suggest that they are superior to κ.

Keywords——QSAR; shape index κ; branching index B.I.; molecular connectivity χ; substituent entropy constant σ_{s°; descriptor μ^2/α; substituted phenyl glycoside; concanavalin A; topological index

Topological indices, such as branching index (B.I.),[1] molecular connectivity χ[2] and shape index κ,[3] have been utilized for many analyses of quantitative structure–activity relationships (QSAR).[4] In our previous communication,[5] it was shown that B.I. and χ are correlated satisfactorily with the absolute entropy S°_{298}(g),[6] and it was suggested that they are originally "non-energetic." The B.I. proposed in the first place by Randic for C_2–C_8 alkanes are based on the counts of C–C bonds. Next, χ's were applied to compounds carrying a functional group (*e.g.* alcohols, amines, *etc.*), but they were originally confined to C_{sp^3}, and the same value as for a C atom is provided for a heteroatom (*e.g.* $^1\chi=1.000$ for CH_3CH_3, CH_3OH, CH_3NH_2, *etc.*). Now, the new descriptor $^n\kappa$ ($n=1,2,3$), having a correction for heteroatoms based on the covalent radius at the valence state, is considered to be more suitable for heteroatoms than χ. The topological index depends originally on the structure of a molecule, and these three topological indices—B.I., $^1\chi$ and $^1\kappa$—should be correlated with one another in aliphatic and aromatic hydrocarbons.

In this work, the relation between σ_{s° and $^n\kappa$ ($n=1,2,3$) has been examined from QSAR examples.

Experimental

Absolute Entropy S°_{298}(g)[6] and Substituent Entropy Constant σ_{s°[7]——The values of S°_{298}(g) were taken from the literature,[6] and the substituent entropy constant σ_{S° is defined by Eq. 1;

$$\sigma_{S^\circ}=\log[S^\circ_{298}(\mathrm{g}),\mathrm{B}/S^\circ_{298}(\mathrm{g}),\mathrm{A}] \qquad (1)$$

where A means methane or benzene, and B means substituted molecules.

Dipole Moment——The values of dipole moments μ [debye] were taken from a data book.[8]

Polarizability——The value of polarizability α [$\times 10^{-24}\,\mathrm{cm}^3$] is given by the Clausius–Mossotti equation as Eq. 2;

$$\alpha=[3(n^2-1)/(n^2+2)\times M/d]/4\pi N\,[10^{-24}\,\mathrm{cm}^3] \qquad (2)$$

where n = refractive index given by $\mathrm{Na_D}$ line at 20 °C, M = molecular weight, d = density at 20 °C, and N = Avogadro's number.

Substituent Constant——The data are cited from reference 9.

Molecular Connectivity $^1\chi$[2]——Kier's report should be referred to for the definition of the descriptor $^1\chi$.

Numerical Treatment——QSAR analyses were carried out on an NEC PC-9801 personal computer utilizing multiregression analysis program MR compiled in the multivariate analysis program package MVA.[10] The standard deviation (S.D.) is given by S.D. $=[S_{\mathrm{se}}/(n-k-1)],^{1/2}$ where n and k denote the numbers of observations and variables, and S_{se} denotes the sum of squares of the residuals. The values in parentheses in the regression equations denote the 95% confidence levels.

Shape Indices $^n\kappa$ ($n=1,2,3$)——Kier's original report should be referred to for the definition of these descriptors.

Results and Discussion

For the 21 congeners in Table I, the following two regression equations were obtained;

$$^1\kappa=2.052(0.111)\mathrm{B.I.}+0.138(2.870) \qquad (3)$$
$$n=17\ \text{(alkanes)},\quad r=0.995,\quad F=1564,\quad \mathrm{S.D.}=0.241$$

$$^1\kappa=2.053(0.091)^1\chi+0.134(2.349) \qquad (4)$$
$$n=21,\quad r=0.996,\quad F=2215,\quad \mathrm{S.D.}=0.215$$

These three kinds of descriptors belong to the category of the entropy of information, but not to the thermodynamic entropy.

Correlations of S°_{298}(g) or σ_{S° and $^1\kappa$ in Alkanes and Alkylbenzenes

For alkanes and alkylbenzenes, the correlations of S°_{298}(g) or σ_{S° and $^1\kappa$ are expressed by Eqs. 5a, b;

TABLE I. Values of S°_{298}(g) [e.u.], Shape Index $^1\kappa$, Molecular Connectivity $^1\chi$ and Branching Index (B.I.) of Alkanes and Alkylbenzenes

	S°_{298}(g)	$^1\kappa$[3]	$^1\chi$[2]	σ_{S°	B.I.
Ethane	54.85	2.000	1.000	0.091	1.000
Propane	64.51	3.000	1.414	0.161	1.414
Butane	74.21	4.000	1.914	0.222	1.914
Pentane	83.40	5.000	2.414	0.273	2.414
Hexane	92.83	6.000	2.914	0.319	2.914
Heptane	102.27	7.000	3.414	0.361	3.414
Octane	111.55	8.000	3.914	0.399	3.914
Nonane	120.86	9.000	4.414	0.434	4.414
Decane	130.17	10.000	4.914	0.466	4.914
Undecane	139.48	11.000	5.414	0.496	5.414
3-Et hexane	109.51	8.000	3.851	0.391	3.851
4-Me heptane	108.35	8.000	3.808	0.386	3.808
3-Me heptane	110.32	8.000	3.808	0.394	3.808
2-Me heptane	108.81	8.000	3.770	0.388	3.770
3,3-Me, Et pentane	103.48	8.000	3.682	0.366	3.682
2,3-$\mathrm{Me_2}$ hexane	106.11	8.000	3.621	0.377	3.621
2,2-$\mathrm{Me_2}$ hexane	103.06	8.000	3.561	0.365	3.561
Benzene	64.34	4.167	1.999	0.156	—
Me benzene	76.64	5.143	2.410	0.236	—
Et benzene	86.15	6.125	2.906	0.287	—
n-Pr benzene	95.76	7.111	3.411	0.333	—

$$S^{\circ}_{298}(\mathrm{g})=9.491(0.800)^{1}\kappa+32.584(0.176) \qquad (5a)$$
$$n=21,\quad r=0.985,\quad F=616.9,\quad \mathrm{S.D.}=3.891$$

$$\sigma_{s}=0.046(0.004)^{1}\kappa+0.017(1.761) \qquad (5b)$$
$$n=21,\quad r=0.982,\quad F=524.5,\quad \mathrm{S.D.}=0.020$$

In our previous report,[5] B.I. and $^{1}\chi$ were correlated better to $S^{\circ}_{298}(\mathrm{g})$ than to $\sigma_{S^{\circ}}$, as given by Eqs. 6a, b and 7a, b;

$$S^{\circ}_{298}(\mathrm{g})=18.502(0.670)\mathrm{B.I.}+38.020(0.006) \qquad (6a)$$
$$n=39,\quad r=0.994,\quad F=3232.9,\quad \mathrm{S.D.}=1.468$$

$$\sigma_{S^{\circ}}=0.094(0.005)\mathrm{B.I.}+0.033(0.458) \qquad (6b)$$
$$n=39,\quad r=0.989,\quad F=1591.8,\quad \mathrm{S.D.}=0.011$$

For alkanes, alkylbenzenes, aliphatic alcohols, amines, ethers, ketones, *etc.*;

$$S^{\circ}_{298}(\mathrm{g})=18.540(1.028)^{1}\chi+38.772(0.084) \qquad (7a)$$
$$n=77,\quad r=0.972,\quad F=1290.5,\quad \mathrm{S.D.}=0.032$$

$$\sigma_{S^{\circ}}=0.071(0.005)^{1}\chi+0.097(0.171) \qquad (7b)$$
$$n=77,\quad r=0.953,\quad F=733.8,\quad \mathrm{S.D.}=0.032$$

As shown in the previous report,[11] E_{dis}, afforded a linear relation with $\sigma_{S^{\circ}}$, but a slight curve with $S^{\circ}_{298}(\mathrm{g})$. Equations 5a, b, 6a, b and 7a, b suggest that the correlations are attributed to the structure of the free molecule, and the above three kinds of topological indices do not directly reflect molecular interaction.

Binding of *p*-Phenyl Glycosides to Concanavalin A[3c]

The data summarized in Table II afforded Eq. 8a, b;

$$\log M_{50}=\underset{(0.33)}{-1.91}\sigma_{S^{\circ}}\underset{(0.16)}{-0.27}\sigma_{\pi}^{+}\underset{(0.19)}{+0.20}\sigma_{\pi}^{-}\underset{(0.02)}{-2.23} \qquad (8a)$$
$$r=0.956,\quad F=53.1,\quad \mathrm{S.D.}=0.04$$

TABLE II. Binding of *p*-Phenyl Glycosides and Concanavalin A

	$\log M_{50}$	$\sigma_{S^{\circ}}$	σ_{π}[9]	$^{2}\kappa_{\alpha}$[3c]
OH	−2.32	0.069	−0.340	1.756
OMe	−2.40	0.127	−0.281	2.478
OEt	−2.46	0.170	−0.281	3.254
OBu	−2.59	0.230	−0.281	4.914
Et	−2.46	0.127	−0.069	2.508
Pr	−2.51	0.173	−0.069	3.286
sec-Bu	−2.58	0.202	−0.051	3.404
tert-Bu	−2.63	0.173	−0.034	2.452
H	−2.23	0	0	1.606
F	−2.26	0.051	−0.118	1.735
Cl	−2.36	0.068	−0.070	1.987
Br	−2.38	0.083	−0.061	2.124
I	−2.44	0.095	−0.068	2.307
COMe	−2.50	0.142	0.266	2.440
COEt	−2.56	0.200	0.266	3.158
CN	−2.35	0.077	0.226	2.131
NO_2	−2.34	0.115	0.332	2.259
NH_2	−2.25	0.074	−0.460	1.756
Me	−2.29	0.076	−0.078	1.783

$$\log M_{50} = -1.86(0.36)\sigma_{S^\circ} - 0.19(0.16)\sigma_\pi^+ - 2.22(0.02) \qquad (8b)$$
$$r=0.942, \quad F=62.6, \quad \text{S.D.}=0.04$$

where the two kinds of descriptors σ_{S° and σ_π^+ are orthogonal ($r=-0.018$).

After regression analysis using the descriptor ${}^2\kappa_\alpha$, the 19 data in Table II, where the congeners of unknown σ_{S° are omitted, afforded Eq. 9;

$$\log M_{50} = \underset{(0.03)}{0.05}({}^2\kappa_\alpha)^2 - \underset{(0.06)}{0.18}{}^2\kappa_\alpha - \underset{(0.06)}{2.00} \qquad (9)$$
$$r=0.877, \quad F=26.6, \quad \text{S.D.}=0.06$$

where the subscript α means the correction due to the heteroatom. Thus, the regression analyses using the descriptor σ_{S° afforded an excellent result, where the contribution from the dispersion and repulsion forces is dominant, and charge transfer interaction is minor.

In the analysis of the biological response, regarded as the reflection of the free energy change, the descriptor σ_{S° is revealed to be a better choice than the shape index, since the former represents the dispersion and repulsion terms, whereas the shape index ${}^2\kappa_\alpha$ originates from the molecular construction, and lacks energetic character.

References

1) M. Randić, *J. Am. Chem. Soc.*, **97**, 6609 (1975).
2) *a*) L. B. Kier, L. H. Hall, W. J. Murray and M. Randić, *J. Pharm. Soc.*, **64**, 1971 (1975); *b*) L. H. Hall, L. B. Kier and W. J. Murray, *ibid.*, **64**, 1974, 1978 (1975).
3) *a*) L. B. Kier, *Quant. Struct.-Act. Relat.*, **4**, 99 (1985); *b*) *Idem, ibid.*, **5**, 1 (1986); *c*) *Idem, ibid.*, **5**, 7 (1986).
4) *a*) L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research," Academic Press, New York, 1976; *b*) L. H. Hall, L. B. Kier and W. J. Murray, *J. Pharm. Sci.*, **64**, 1974 (1975); *c*) W. J. Murray, L. H. Hall and L. B. Kier, *ibid.*, **64**, 1978 (1975).
5) Y. Sasaki, T. Takagi, H. Kawaki and A. Iwata, *Chem. Pharm. Bull.*, **31**, 330 (1983).
6) *a*) D. R. Stull, E. F. Westrum, Jr. and G. C. Sinke, "The Chemical Thermodynamics of Organic Compounds," John Wiley & Sons, Inc., New York, 1969; *b*) S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw and R. Walsh, *Chem. Rev.*, **69**, 279 (1969).
7) Y. Sasaki, T. Takagi, Y. Yamazato, A. Iwata and H. Kawaki, *Chem. Pharm. Bull.*, **29**, 3073 (1981).
8) A. L. McClellan, "Tables of Experimental Dipole Moments," Freeman, San Francisco, 1963.
9) *a*) Y. Yukawa and Y. Tsuno, *Nippon Kagaku Kaishi*, **86**, 873 (1965); *b*) M. Sawada, M. Ichihara, Y. Yukawa, T. Nakachi and Y. Tsuno, *Bull. Chem. Soc. Jpn.*, **53**, 2055 (1980).
10) T. Takagi, K. Tange, N. Iwata, T. Katayama, H. Izawa, S. Fujii and Y. Sasaki, Proceedings of the 4th Software Conference, Osaka, 1988, p. 285.
11) Y. Sasaki, T. Takagi and H. Kawaki, *Chem. Pharm. Bull.*, **36**, 3743 (1988).