Procedures on the Estimation of the Novel Quantitative Structure–Activity Relationship (QSAR) Descriptor σ_s for Poly-Substituted Benzene Series, and Their Polarizabilities

Yoshio SASAKI,^a Hideko KAWAKI,^b and Tatsuya TAKAGI^{*,a}

Faculty of Pharmaceutical Sciences, Osaka University,^a 1–6 Yamada-Oka, Suita, Osaka 565, Japan and Faculty of Pharmacy, Kinki University,^b 3–4–1 Kowakae, Higashi-Osaka, Osaka 577, Japan. Received June 13, 1990

Procedures for the estimation of the novel quantitative structure–activity relationship (QSAR) descriptor σ_s , representing the contributions from both dispersion and repulsion interactions, have been presented for the poly-substituted benzene derivatives.

Observed values σ_s for $C_6H_{6-n}R_n$ ($R = F, Cl, Me, Et; n = 2–6$) can be given by the equation $\sigma_s(R_n) = a \log n + b$, where the slope a and the intercept b are also linear against $\sigma_s(\text{mono})$. Consequently, the slope a and the intercept b estimated from $\sigma_s(\text{mono})$ and $\log n$ afford σ_s for poly-substituted benzene derivatives having an optional set of substituent groups. The values thus obtained agreed well with the observed ones, and were also found to be linear against $\Sigma\sigma_s(\text{mono})$, where the correction of the number of symmetry should be taken into account.

Furthermore, the net increases of polarizabilities $\Delta\alpha$ for the tetra-, penta-, and hexa-substituted benzene series from benzene reference are approved to be expressed by the linear relation against $\Sigma\Delta\alpha(\text{mono})$.

Keywords quantitative structure–activity relationship; entropy constant σ_s ; poly-substituted benzene derivative; polarizability α ; regression analysis

Introduction

In previous reports^{1a–c)} on the studies of the estimation of our novel quantitative structure–activity relationship (QSAR) descriptors σ_s , the authors determined those of mono-, di-, and tri-substituted benzene derivatives where linear relations among the observed or estimated descriptors σ_s and $\Sigma\sigma_s(\text{mono})$ were acknowledged.

In this work, in order to determine the values of σ_s for the poly-substituted benzene series, the authors have utilized the linear relations between $\sigma_s(R_n)$ ($n = 2–6$, $R = F, Cl, Me, Et$) and $\log n$, where n denotes the number of the substituent. The slope and the intercept determined in the above four series were also found to be linear against $\sigma_s(\text{mono})$. The determination of the two factors written above enabled us to estimate the descriptor $\sigma_s(R_n)$.

In addition, the polarizabilities α for the polysubstituted benzene series have been approved as a linear function of $\Sigma\Delta\alpha(\text{mono})$ from benzene reference.

Experimental

Descriptor σ_s . Observed data for mono- and poly-substituted benzene series were determined from $S_{298}^\circ(g)$ data sources^{2a–e)} by the usual way.

Polarizability α . Polarizability/ 10^{-24} cm^3 was derived from the Clausius–Mossotti's equation $\alpha = n^2 - (1/n^2) + 2 \cdot (M/d) \cdot \{3/(4N)\}$, where refractive index n and density d were cited from the references.^{3a,b)}

Regression Analyses Regression analyses were carried out on NEC PC-9801/M/VX and EPSON PC-286V personal computers using a program package for multi-variate analyses, MVA developed by Takagi *et al.*⁴⁾

Results and Discussion

Descriptor σ_s for $C_6H_{6-n}R_n$ In the first step of this work, the two procedures written below were employed; (1) The parametrization of $\sigma_s(n) = \log a + b \log n + 2.303cn$, supposing the curve-fitting on an equation expressed by $\sigma_s(n) = an^b e^{cn}$, where n means the number of the substituent group. (2) The parametrization by $\sigma_s(n) = a + \log n$, $n \geq 2$, based on the relation $S_{298}^\circ(g)^{1/2} \propto n^{1/2}$, where the number $n = 1–6$ could be regarded as a sort of entropy of information.

But, unfortunately the curve-fitting of process (1) failed. The results reproduced in the higher homologs were

attended by inevitable deviations in the lower homologs, and the reverse situations were also observed. In addition, for process (2), the scattering of the plot is unavoidable in $C_6H_{6-n}X_n$ ($X = F, Cl$) because the slopes of $\sigma_s(n)$ against $\log n$ are gentle.

Under the above evidence, the authors examined the linear relations between $\sigma_s(n)$ and $\log n$ for $n = 2–6$, rejecting the data of the monosubstituted benzene series, because the values of σ_s determined from the observed $S_{298}^\circ(g)$ were already proposed in our previous work.^{1a)}

Plots of the data against the logarithm of the number n of the substituent are linear, and after regression they are formulated as Eqs. 1–4 as given below:

$$\sigma_s(F_n) = 0.1614(\pm 0.0301)\log n + 0.0221(\pm 0.0161) \quad (1)$$

$$n = 11, r = 0.972, s = 0.0069, d = 2.103$$

$$\sigma_s(Cl_n) = 0.2305(\pm 0.0250)\log n + 0.0307(\pm 0.0133) \quad (2)$$

$$n = 11, r = 0.990, s = 0.0057, d = 2.659$$

$$\sigma_s(Me_n) = 0.2375(\pm 0.0229)\log n + 0.0479(\pm 0.0122) \quad (3)$$

$$n = 11, r = 0.992, s = 0.0052, d = 2.416$$

$$\sigma_s(Et_n) = 0.4271(\pm 0.0187)\log n + 0.0786(\pm 0.0100) \quad (4)$$

$$n = 11, r = 0.998, s = 0.0043, d = 2.023$$

TABLE I. Observed Descriptors σ_s (obs) for $C_6H_{6-n}R_n$ ($R = F, Cl, Me, Et; n = 1–6$) and Number n of Substituent R

n	$\log n$	Position	F	Cl	Me	Et
1	0.000	1	0.051	0.066	0.076	0.127
2	0.301	12	0.078	0.103	0.117	0.208
		13	0.076	0.106	0.123	0.213
		14	0.069	0.097	0.117	0.208
3	0.477	123	0.095	0.138	0.155	0.275
		124	0.109	0.145	0.167	0.284
		135	0.088	0.129	0.156	0.276
4	0.602	1234	0.115	0.169	0.196	0.333
		1235	0.122	0.172	0.196	0.338
		1245	0.119	0.165	0.192	0.335
5	0.699		0.144	0.199	0.217	0.381
6	0.778		0.153	0.209	0.225	0.413

TABLE II. Estimated Slopes and Intercepts for Typical Substituent Groups

		Slope	Intercept
1	NO ₂	0.384	0.071
2	CN	0.251	0.042
3	COMe	0.473	0.090
4	COEt	0.606	0.119
5	CF ₃	0.473	0.090
6	COOMe	0.623	0.123
7	COOEt	0.725	0.145
8	F	0.166	0.023
9	Cl	0.217	0.034
10	Br	0.260	0.047
11	I	0.310	0.058
12	Me	0.251	0.042
13	Et	0.425	0.080
14	OMe	0.425	0.080
15	OEt	0.572	0.112
16	OH	0.230	0.039
17	NH ₂	0.240	0.043
18	NMe ₂	0.450	0.088

TABLE III. Descriptors $\sigma_s(\text{obs})$ and Their Related Data for Poly-Substituted Benzene Derivatives

		Observed		Calculated		ΔS°	$\Sigma\sigma_s(\text{mono})$
		σ_s°	S°	σ_s°	S°		
1	1,2,3,4-Me ₄	0.190	99.55	0.191	99.88	-0.33	0.304
2	-Et ₄	0.333	138.45	0.336	139.47	-1.02	0.508
3	-F ₄	0.115	83.45	0.121	85.01	-1.56	0.204
4	-Cl ₄	0.169	94.91	0.169	94.95	-0.04	0.264
5	-(COOMe) ₄	0.498 ^{a)}	202.53 ^{a)}	0.498 ^{b)}	202.53 ^{b)}		0.740
6	-(COOEt) ₄	0.582 ^{a)}	245.74 ^{a)}	0.581 ^{b)}	245.18 ^{b)}		0.860
7	1,2,3,5-Me ₄	0.196	100.99			1.11	
8	-Et ₄	0.338	139.99			0.52	
9	-F ₄	0.122	85.21			0.20	
10	-Cl ₄	0.172	95.50			0.55	
11	-(COOMe) ₄	0.502 ^{a)}	204.40 ^{a)}				
12	-(COOEt) ₄	0.586 ^{a)}	248.02 ^{a)}				
13	1,2,4,5-Me ₄	0.192	100.03			0.45	
14	-Et ₄	0.335	139.03			-0.44	
15	F ₄	0.119	84.62			-0.39	
16	-Cl ₄	0.165	94.03			-0.92	
17	-(COOMe) ₄	0.499 ^{a)}	202.99 ^{a)}				
18	-(COOEt) ₄	0.584 ^{a)}	246.88 ^{a)}				
19	-Me ₅	0.217	106.09	0.214	105.31	0.78	0.380
20	-Et ₅	0.381	154.84	0.377	153.28	1.56	0.635
21	-F ₅	0.144	89.64	0.137	88.20	1.44	0.255
22	-Cl ₅	0.199	101.74	0.192	100.07	1.67	0.330
23	-(COOMe) ₅	0.558 ^{a)}	232.53 ^{a)}	0.558 ^{b)}	232.53 ^{b)}		0.925
24	-(COOEt) ₅	0.652 ^{a)}	288.72 ^{a)}	0.652 ^{b)}	288.72 ^{b)}		1.075
25	-Me ₆	0.225	108.12	0.233	109.95	-1.83	0.456
26	-Et ₆	0.413	166.62	0.411	165.72	0.90	0.762
27	-F ₆	0.153	91.54	0.150	91.30	0.24	0.306
28	-Cl ₆	0.209	104.13	0.210	102.68	1.45	0.396
29	-(COOMe) ₆	0.608	260.90 ^{a)}	0.608 ^{b)}	260.90 ^{b)}		1.110
30	-(COOEt) ₆	0.709	329.22 ^{a)}	0.709 ^{b)}	329.22 ^{b)}		1.290
31	-F ₅ , Me	0.186	98.74	0.196	101.00	-3.26	0.331
32	-F ₅ , Cl	0.182	97.83	0.191	99.77	-1.94	0.321
33	-F ₅ , Br	0.195	100.80	0.198	101.62	-0.82	0.336
34	-F ₅ , I	0.204	102.92	0.205	103.24	-0.32	0.349

a) The numerical values were obtained from Eqs. 7—11. b) The numerical values were obtained from Eqs. 5—6.

where the four kinds of slopes and intercepts thus determined are revealed to be linear against $\sigma_s(\text{mono})$ as below:

TABLE IV. Polarizabilities $\alpha/10^{-24} \text{ cm}^3$ and Their Related Quantities for Poly-Substituted Benzene Derivatives

	$\Delta\alpha(\text{obs})$	$\Sigma\Delta\alpha(\text{mono})$
1,2,3,4-Tetra-		
1 1,2,3,4-Me ₄	7.491	7.716
2 1,2,3,4-Et ₄	15.152	15.200
3 2-Et-1,3,4-Me ₃	9.366	9.587
4 4-Et-1,2,3-Me ₃	9.366	9.587
1,2,4,6-Tetra-		
1 1-COCl-2,4,6-Me ₃	9.896	10.230
2 1,2,4,6-Me ₄	7.580	7.716
3 2-Et-1,3,5-Me ₃	9.441	9.589
4 6-Et-1,2,4-Me ₃	9.435	9.589
5 5-Et-1,2,3-Me ₃	9.455	9.589
6 1-n-Pr-2,4,6-Me ₃	11.285	11.442
7 1-Cl-2,4,6-Me ₃	7.684	7.848
8 1-Me-2-NO ₂ -4-iso-Pr-6-Cl	11.670	12.146
9 1-F-2,4,6-Cl ₃	5.906	5.875
1,2,4,5-Tetra-		
1 1-iso-Pr-2,4,5-Me ₃	11.367	11.432
2 1-CH=CH ₂ -2,4,5-Me ₃	9.666	9.843
3 1-Et-2,4,5-Me ₃	10.634	9.587
4 1-n-Pr-2,4,5-Me ₃	11.293	11.442
5 1,2,4,5-F ₄	-0.105	-0.176
6 1,2,4,5-Et ₄	15.103	15.200
7 1-Br-n-Pr ₃	20.057	20.057
8 1-NO ₂ -2,4,5-n-Pr ₃	19.430	19.564
9 1-Me-2-F-4-NO ₂ -5-Cl	6.372	6.457
Penta-		
1 4-Br-2-C=CH-1,3,5-Me ₃	12.212	12.573
2 3-CH ₂ Cl-1,2,4,6-Me ₄	11.533	11.628
3 Me ₅	9.358	9.645
4 3-C=CH-1,2,4,6-Me ₄	10.932	11.412
Hexa-		
1 F ₆	0.119	-0.264
2 F ₅ Cl	2.728	1.753
3 6-CF ₃ -2,3,4,5-F ₄ -1-Cl	4.031	3.616
4 F ₅ CF ₃	2.260	1.599
5 Cl ₅ -CH=CH ₂	12.791	13.921

$$\text{slope} = 3.4109(\pm 1.0274)\sigma_s(\text{mono}) - 0.0080(\pm 0.0087) \quad (5)$$

$$n=4, r=0.995, s=0.0136, d=3.212$$

$$\text{intercept} = 0.7454(\pm 0.3885)\sigma_s(\text{mono}) - 0.0148(\pm 0.0330) \quad (6)$$

$$n=4, r=0.986, s=0.0052, d=2.929$$

where script *d* means the Durbin-Watson ratio.⁵⁾ The slope and the intercept determined by the above procedures are all summarized in Table II, including the typical substituent groups. The data of observed $\sigma_s(\text{obs})$, as well as those of the calculated $\sigma_s(\text{cal})$ given by Eqs. 5—6, for the poly-substituted benzene series ($n=4-6$), including the additional data of CO₂R (R=Me, Et) are arranged in Table III, where the agreements within ± 2 e.u. are accepted.

Furthermore, in the next step, we have expressed the values $\sigma_s(\text{obs})$ as a linear function of $\Sigma\sigma_s(\text{mono})$ as an independent variable, and regression analyses afforded Eqs. 7—11 as given below:

$$\sigma_s(\text{obs})/1234 = 0.7043(\pm 0.1128)\Sigma\sigma_s(\text{mono}) - 0.0236(\pm 0.0384) \quad (7)$$

$$n=4, r=0.9986, s=0.0060, F=720.6$$

$$\sigma_s(\text{obs})/1235 = 0.7017(\pm 0.0757)\Sigma\sigma_s(\text{mono}) - 0.0175(\pm 0.0257) \quad (8)$$

$$n=4, r=0.9994, s=0.0040, F=1590.0$$

$$\sigma_s(\text{obs})/1245 = 0.7063(\pm 0.0356)\Sigma\sigma_s(\text{mono}) - 0.0233(\pm 0.0121) \quad (9)$$

$$n=4, r=0.9999, s=0.0019, F=7284.7$$

$$\sigma_s(\text{obs})/n=5 = 0.6178(\pm 0.0993)\Sigma\sigma_s(\text{mono}) - 0.0119(\pm 0.0422) \quad (10)$$

$$n=4, r=0.9986, s=0.0066, F=717.4$$

$$\sigma_s(\text{obs})/n=6 = 0.5696(\pm 0.1209)\Sigma\sigma_s(\text{mono}) - 0.0234(\pm 0.0616) \quad (11)$$

They are, unfortunately, concerned with the homologs having the same substituent groups, and, when the loss of the symmetry number results from introducing different kinds of substituents, it is desirable to take into account correction due to the symmetry loss. The evidence is provided by the data of congeners 31—34, *e.g.* $\text{C}_6\text{F}_5\text{R}$ ($\text{R}=\text{Me}, \text{Cl}, \text{Br}, \text{I}$)^{2b-e)} arranged in Table III, where the correction of 5Rln 2 e.u. is required from the reference value because the symmetry numbers inherent in $\Sigma\sigma_s(\text{mono})$ and $\text{C}_6\text{F}_5\text{R}$ are 6 and 1, respectively.

Polarizability α and Poly-Substituted Benzene Series In this section, we have determined the net increase of polarizability $\Delta\alpha$ for the poly-substituted benzene series^{3a,b)} from benzene reference, as well as the sum of the excess amount of polarizability $\Sigma\Delta\alpha(\text{mono})$ determined individually from mono-substituted benzene derivatives,⁶⁾ respectively.

The results are all summarized in Table IV, where the two kinds of data are correlated successfully as given by Eqs. 12—16, and from which the values of $\Delta\alpha$ for the poly-substituted benzene series having optical kinds of substituent groups could be realized.

$$\Delta\alpha/1234 = 1.0260(\pm 0.0206)\Sigma\Delta\alpha(\text{mono}) - 0.4528(\pm 0.2240) \quad (12)$$

$$n=4, r=1.0000, s=0.0268, F=46065$$

$$\Delta\alpha/1235 = 0.9428(\pm 0.0404)\Sigma\Delta\alpha(\text{mono}) + 0.3464(\pm 0.3845) \quad (13)$$

$$n=9, r=0.9989, s=0.0938, F=3042.4$$

$$\Delta\alpha/1245 = 0.9890(\pm 0.0532)\Sigma\Delta\alpha(\text{mono}) + 0.1724(\pm 0.6892) \quad (14)$$

$$n=9, r=0.9982, s=0.4028, F=1929.5$$

$$\Delta\alpha/n=5 = 0.9884(\pm 0.4004)\Sigma\Delta\alpha(\text{mono}) - 0.1750(\pm 4.5504) \quad (15)$$

$$n=4, r=0.9913, s=0.1970, F=112.80$$

$$\Delta\alpha/n=6 = 0.8671(\pm 0.1018)\Sigma\Delta\alpha(\text{mono}) + 0.8090(\pm 0.6637) \quad (16)$$

$$n=5, r=0.9980, s=0.3612, F=734.8$$

Acknowledgement Financial assistance from the Fund of Hoan-sha is gratefully acknowledged.

References and Notes

- 1) a) Y. Sasaki, T. Takagi, and H. Kawaki, *Chem. Pharm. Bull.*, **36**, 3743 (1988); b) H. Kawaki, Y. Sasaki, T. Takagi, S. Fujii, and F. Masuda, *ibid.*, **37**, 3268 (1989); c) Y. Sasaki, H. Kawaki, and T. Takagi, *ibid.*, **38**, 2344 (1990).
- 2) a) D. R. Stull, E. F. Westrum, Jr., and G. C. Sinke, "The Chemical Thermodynamics of Organic Compounds," Wiley, New York, 1969; b) J. B. Butler and J. Liemezs, *J. Chem. Eng. DATA*, **14**, 335 (1969); c) J. F. Counsell, J. L. Hales, and J. F. Martin, *J. Chem. Soc. (A)*, **1968**, 2042; d) R. J. L. Andon, J. F. Counsell, J. L. Hales, E. B. Lees, and J. F. Martin, *ibid.*, **1968**, 2357; e) J. F. Counsell, J. L. Hales, E. B. Lees, and J. F. Martin, *ibid.*, **1968**, 2994.
- 3) a) Zvi Rappoport, "Handbook of Tables for Organic Compound Identification," 3rd Ed., Chemical Rubber Co., Ohio, 1969; b) Beilstein's "Handbuch der Organischen Chemie," Drittes Ergänzungswerke, Band. V, Springer Verlag, Berlin, 1963—1965.
- 4) T. Takagi, K. Tange, N. Iwata, Y. Shindo, A. Iwata, T. Katayama, H. Izaki, S. Fujii, and Y. Sasaki, "Proceedings of the 4th Software Conference," Osaka, 1988, p. 285.
- 5) J. Durbin and G. S. Watson, *Biometrika*, **58**, 1 (1971).
- 6) H. Kawaki, F. Masuda, and Y. Sasaki, *Chem. Pharm. Bull.*, **36**, 4814 (1988).