Chem. Pharm. Bull. 40(3) 565—569 (1992)

On the Estimation of the Quantitative Structure–Activity Relationships Descriptor σ_{s° for Aliphatic Compound

Yoshio SASAKI,[a] Tatsuya TAKAGI,*[,a] and Hideko KAWAKI[b]

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 July 11, 1991*

Procedures for the estimation of the novel quantitative structure–activity relationships (QSAR) descriptors σ_{s°, representing the contributions from both dispersion and repulsion interactions, have been presented for the substituted methane series. After correcting part of the symmetry, the rotation isomer or racemate, observed values σ_{s°(obs) for $CH_2R_AR_B$, $CHR_AR_BR_C$, and $CR_AR_BR_CR_D$ can be given by the general equation $\sigma_{s^\circ}(\text{obs}) = a\Sigma\sigma_{s^\circ}(\text{mono}) + b$, where σ_{s°(mono) means the descriptor of the MeR series. The values estimated by this equation agreed with those of the observed ones, and several examples of the large molecule are submitted for the determination of the optimum value.

Furthermore, the hydrophobic substituent constants π of typical α-amino acids are revealed to be expressed by the linear combination of the descriptor σ_{s° based on the gas phase and indicator variable, implying the averaged number of hydration.

Keywords quantitative structure–activity relationship; substituent entropy constant σ_{s°; substituted methane derivative; regression analysis; amino acid

Introduction

In our previous report,[1] the method for the estimation of the quantitative structure–activity relationships (QSAR) descriptor σ_{s° of polysubstituted benzene derivatives having an optional set of substituent groups has been confirmed by means of the statistical way, and the results are fruitful in the evaluation of the phenomena inherent in the van der Waals force, including gas-liquid chromatography (GLC),[2] *etc.*

In this work, the authors focused their attention on the estimation of the descriptor σ_{s° for substituted aliphatic substrates $CH_2R_AR_B$, $CHR_AR_BR_C$, and $CR_AR_BR_CR_D$. The successful result after regression analyses enabled us to predict the descriptor σ_{s° for the aliphatic substrate having an optional set of substituent groups.

Experimental

Absolute Entropy S°_{298}(g)/e.u. Data are all cited from the references.[3,4]

Regression Analysis Regression analyses are 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.*[5]

Definition of the Descriptor σ_{s°(obs) for Substituted Methane Derivatives The observed descriptors σ_{s°(obs) for MeR,[6] $CH_2R_AR_B$, $CHR_AR_BR_C$ and $CR_AR_BR_CR_D$ are defined as follows; $\sigma_{s^\circ}(\text{obs}) = \log\{S^\circ_{298}(\text{g})(\text{obs})/S^\circ_{298}(CH_4)\}$, where $S^\circ_{298}(\text{g})(CH_4) = 44.52$ e.u. is used.

Results and Discussion

Correlations between σ_{s° and $\Sigma\sigma_{s^\circ}$(mono) in Substituted Methanes The successful results[1] on the estimation of the descriptor σ_{s° for polysubstituted benzene derivatives having an optional set of substituent as a function of $\Sigma\sigma_{s^\circ}$(mono) have prompted us to develop the strategy in the field of aliphatic substrate.

Followed by the definition in our previous work,[6] the presentation of the QSAR descriptor σ_{s°(mono) or mono-substituted methane series is given as below;

$$\sigma_{s^\circ}(\text{mono}) = \log\{S^\circ_{298}(\text{g})(\text{MeR})/S^\circ_{298}(\text{g})(CH_4)\}$$

where CH_4 is used as reference.

In the first step, the values of σ_{s°(obs) derived from the observed absolute entropies S°_{298}(g) and calculated $\Sigma\sigma_{s^\circ}$(mono) for the $CH_2R_AR_B$($R_A = R_B$, $R_A \neq R_B$), $CHR_AR_BR_C$($R_A = R_B = R_C$, $R_A = R_B \neq R_C$, $R_A \neq R_B \neq R_C$), and $CR_AR_BR_CR_D$ ($R_A = R_B = R_C = R_D$, $R_A = R_B = R_C \neq R_D$, $R_A = R_B \neq R_C \neq R_D$, $R_A = R_B \neq R_C = R_D$) series are arranged in Tables I, II, and III, where additional corrections for σ_{s°(obs) are carried out, namely, $n \times \boldsymbol{R}\ln 2$ in the presence of the plane of symmetry, $\boldsymbol{R}\ln 3$ for *trans* and *gauche* isomer and $\boldsymbol{R}\ln 2$ for optical antipode. The abbreviations n and $\boldsymbol{R}$ denote the number of symmetry and gas constant, respectively. The observed data and those having the correction term

TABLE I. $CH_2R_AR_B$ and Their Descriptors σ_{s°

	S°_{298}(obs)	σ_{s°(obs)	Correction	σ_{s°(cor)	σ_{s°(mono)	σ_{s°(cal)
1 CH_2F_2[a)]	58.94	0.122	$2 \times \boldsymbol{R}\ln 2$	0.142	0.156	0.148
2 CH_2Cl_2	64.59	0.162	$2 \times$	0.180	0.198	0.179
3 CH_2Br_2	70.08	0.197	$2 \times$	0.214	0.244	0.213
4 CH_2I_2	73.88	0.220	$2 \times$	0.236	0.270	0.233
5 $(Me)_2CH_2$	64.51	0.161			0.182	0.167
6 $(Et)_2CH_2$	83.40	0.273			0.322	0.271
7 $(n\text{-}Pr)_2CH_2$	102.27	0.361			0.442	0.361
8 CH_2BrCl	68.67	0.188	$1 \times$	0.197	0.221	0.196
9 CH_2BrF	65.97	0.171	$1 \times$	0.180	0.200	0.181
10 CH_2BrI	73.49	0.218	$1 \times$	0.226	0.257	0.223
11 CH_2ClF	63.16	0.152	$1 \times$	0.161	0.177	0.163
12 CH_2ClI	70.78	0.201	$1 \times$	0.210	0.234	0.206
13 $MeCH_2Br$	68.71	0.188			0.213	0.190
14 $MeCH_2Cl$	65.93	0.171			0.190	0.173
15 $MeCH_2F$	63.32	0.153			0.169	0.157
16 $MeCH_2NH_2$	68.08	0.184			0.206	0.185
17 $MeCH_2NO_2$	75.39	0.229			0.260	0.225
18 $MeCH_2OH$	67.54	0.181			0.201	0.181
19 $MeCH_2SH$	70.77	0.210			0.227	0.201
20 $MeCH_2CN$	68.50	0.187			0.211	0.189
21 $MeCH_2CHO$	72.83	0.214			0.243	0.213
22 $(MeCH_2)_2$	74.12	0.221			0.252	0.219
23 $(ClCH_2)_2$	73.66	0.219	$1 \times \boldsymbol{R}\ln 3$	0.231	0.271	0.233
24 $(BrCH_2)_2$	78.81	0.248	$1 \times$	0.260	0.310	0.262
25 $(ICH_2)_2$	83.30	0.272	$1 \times$	0.283	0.336	0.282
26 $(HOCH_2)_2$	77.33	0.240	$1 \times$	0.252	0.292	0.249
27 $(NCCH_2)_2$	79.04	0.249	$1 \times$	0.261	0.309	0.262
28 $(tert\text{-}BuCH_2)_2$	112.35	0.402			0.494	0.400
29 $(n\text{-Pentyl})_2S$	144.45	0.513			0.664	0.526
30 $(n\text{-Bu})_2S$	125.84	0.451			0.566	0.453

a) MeF, having $S^\circ_{298}(\text{g}) = 53.25$ e.u., gives $\sigma_{s^\circ}(\text{mono}) = 0.078$, and for CH_2F_2, we get $\Sigma\sigma_{s^\circ}(\text{mono}) = 0.156$. Using Eq. 1, CH_2F_2 belonged to $CH_2R_AR_B$, affords $\sigma_{s^\circ}(\text{cal}) = 0.148$. On the other hand, CH_2F_2, having $S^\circ_{298}(\text{g})(\text{obs}) = 58.94$ e.u. and $\sigma_{s^\circ}(\text{obs}) = 0.122$, affords, after correcting the symmetry part $2 \times \boldsymbol{R}\ln 2 = 2.75$ e.u., $S^\circ_{298}(\text{g})(\text{corr.}) = 58.94 + 2.75 = 61.69$ e.u. and $\sigma_{s^\circ}(\text{corr.}) = 0.142$.

TABLE II. $CHR_AR_BR_C$ and Their Descriptors σ_{s°

		S°_{298}(g)(obs)	σ_{s°(obs)	Correction	σ_{s°(cor)	$\Sigma\sigma_{s^\circ}$(mono)	σ_{s°(cal)
1	CHF_3	62.04	0.144	$3\times R\ln 2$	0.172	0.234	0.182
2	$CHCl_3$	70.66	0.201	3×	0.225	0.297	0.224
3	$CHBr_3$	79.03	0.249	3×	0.271	0.366	0.269
4	CHI_3	85.00	0.281	3×	0.301	0.405	0.295
5	$CHMe_3$	70.42	0.199			0.273	0.208
6	$CHEt_3$	98.35	0.344			0.483	0.346
7	$CHBrCl_2$	75.56	0.230	1×	0.238	0.326	0.243
8	$CHBrF_2$	70.51	0.200	1×	0.208	0.278	0.211
9	$CHBr_2Cl$	78.31	0.245	1×	0.253	0.343	0.254
10	$CHBr_2F$	75.70	0.231	1×	0.238	0.322	0.240
11	$CHClF_2$	67.13	0.178	1×	0.187	0.255	0.196
12	$CHCl_2F$	69.99	0.196	1×	0.205	0.276	0.210
13	Me_2CHOH	74.03	0.221			0.292	0.220
14	Me_2CHSH	77.51	0.241			0.318	0.238
15	Me_2CHNO_2	83.10	0.271			0.351	0.259
16	Me_2CHCl	72.70	0.213			0.281	0.213
17	Me_2CHBr	85.53	0.230			0.304	0.228
18	Me_2CHF	69.82	0.195			0.260	0.199
19	Me_2CHI	77.55	0.241			0.317	0.237
20	Me_2CHCN	74.88	0.226			0.302	0.227
21	MeO–iso-Pr	80.86	0.259			0.338	0.251
22	MeS–iso-Pr	85.87	0.285			0.368	0.270
23	*n*-$PrCHEt_2$	109.51	0.391			0.543	0.386
24	*n*-Pr_2CHEt	118.52	0.425			0.603	0.425
25	$CHBrClF$[a]	72.88	0.214	−1×	0.206	0.298	0.224
26	MeCHClEt[a]	85.94	0.286	−1×	0.279	0.351	0.259
27	MeCHOHEt[a]	85.81	0.285	−1×	0.278	0.362	0.267
28	MeCHSHEt[a]	87.65	0.294	−1×	0.287	0.388	0.284
29	$MeCHNH_2Et$[a]	83.90	0.275	−1×	0.268	0.367	0.270
30	MeEtCH–*n*-Pr[a]	101.37	0.357	−1×	0.351	0.473	0.340
31	$(n\text{-Pr})_3CH$	125.56	0.450		0.663	0.465	
32	$(n\text{-Pr})_2CH$–iso-Pr	124.63	0.447		0.641	0.450	
33	*n*-PrEtCH–*n*-Bu[a]	129.12	0.462	−1×	0.458	0.655	0.459

a) Racemic compound.

afford excellent linear relations with $\Sigma\sigma_{s^\circ}$(mono) as given below;

$$\sigma_{s^\circ}(CH_2R_AR_B)=0.745(0.011)\Sigma\sigma_{s^\circ}(\text{mono})+0.032(0.003) \quad (1)$$
$$n=30,\ r=0.999,\ s=0.003$$

$$\sigma_{s^\circ}(CHR_AR_BR_C)=0.658(0.025)\Sigma\sigma_{s^\circ}(\text{mono})+0.028(0.010) \quad (2)$$
$$n=32,\ r=0.995,\ s=0.008$$

$$\sigma_{s^\circ}(CR_AR_BR_CR_D)=0.684(0.022)\Sigma\sigma_{s^\circ}(\text{mono})-0.013(0.010) \quad (3)$$
$$n=40,\ r=0.995,\ s=0.005$$

And, the calculated values of σ_{s°(cal) given by Eqs. 1—3 are all aligned in the extreme right of the Tables.

Using Eq. 1, we are able to estimate the descriptor for important $CH_2R_AR_B$ type substrate, as given in Table IVa.

Optimum of Descriptor σ_{s°(cal) and $\Sigma\sigma_{s^\circ}$(mono) as Fragment Sum In contrast to benzene derivatives, aliphatic substrate composed of a variety of binding fragments does not specify the optimal $\Sigma\sigma_{s^\circ}$(mono). This situation is approved for several kinds of examples given in Chart 1, where the fragment tentatively specified by the optional reference position affords good agreement among observed and calculated descriptors, and the following conclusions are obtained;

1) The average of $\Sigma\sigma_{s^\circ}$(mono) is desirable.

2) The disparty of the size of the fragment should be kept out.

3) Under conditions 1) and 2), an optional choice of fragment affords that the deviation not exceed 5% from S°_{298}(g)(obs).

From the above view point, the authors selected and estimated the optimal $\Sigma\sigma_{s^\circ}$(mono) and σ_{s°(cal) for substrates summarized in Table IVb

Amino Acids and Their Descriptors σ_{s° Successful results

TABLE III. $CR_AR_BR_CR_D$ and Their Descriptors σ_{s°

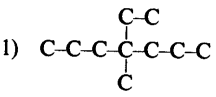
		S°_{298}(obs)	σ_{s°(obs)	Correction	σ_{s°(cor)	$\Sigma\sigma_{s^\circ}$(mono)	σ_{s°(cal)
1	ClF_2CCN	76.08	0.233	$1\times R\ln 2$	0.241	0.375	0.243
2	Cl_2FCCN	79.33	0.251	1×	0.258	0.396	0.257
3	Cl_3CCN	79.38	0.251	3×	0.273	0.417	0.272
4	F_3CCN	71.27	0.204	3×	0.229	0.354	0.229
5	Me_4C	73.23	0.216	6×	0.263	0.364	0.236
6	Et_4C	110.31	0.394	6×	0.425	0.644	0.427
7	CF_4	62.50	0.147	6×	0.201	0.312	0.200
8	CCl_4	74.12	0.221	6×	0.267	0.396	0.257
9	CBr_4	85.53	0.284	6×	0.324	0.488	0.320
10	CI_4	93.60	0.323	6×	0.359	0.540	0.356
11	$CBrClF_2$	76.14	0.233	1×	0.241	0.377	0.244
12	$CBrCl_2F$	78.87	0.248	1×	0.256	0.398	0.259
13	$CBrCl_3$	79.55	0.252	3×	0.274	0.419	0.273
14	$CBrF_3$	71.19	0.204	3×	0.228	0.356	0.230
15	CBr_2ClF	81.89	0.265	1×	0.272	0.421	0.275
16	CBr_2Cl_2	83.23	0.272	2×	0.286	0.442	0.289
17	CBr_2F_2	77.66	0.242	2×	0.257	0.400	0.260
18	CBr_3Cl	85.36	0.283	3×	0.303	0.465	0.305
19	CBr_3F	82.65	0.269	3×	0.290	0.444	0.290
20	CCl_3F	68.28	0.186	3×	0.211	0.333	0.214
21	CCl_2F_2	71.91	0.208	2×	0.225	0.354	0.229
22	$CClF_3$	74.13	0.221	3×	0.245	0.375	0.243
23	CF_3I	73.50	0.218	3×	0.241	0.369	0.239
24	CF_2I_2	82.78	0.269	2×	0.284	0.426	0.278
25	CCl_3I	81.34	0.262	3×	0.282	0.432	0.282
26	Me_3CCl	77.00	0.238			0.372	0.241
27	Me_2CCl_2	77.92	0.243			0.380	0.247
28	Me_3CBr	79.34	0.251			0.395	0.257
29	Me_3CI	81.79	0.264			0.408	0.266
30	Me_3COH	77.98	0.243			0.383	0.249
31	Me_3CSH	80.79	0.259			0.409	0.266
32	$EtMe_2COH$	87.68	0.294			0.453	0.296
33	$EtMe_2CSH$	92.48	0.317			0.479	0.314
34	$EtCMe_3$	85.62	0.284			0.434	0.283
35	*n*-$PrCMe_3$	93.90	0.324			0.494	0.324
36	$Me_2C(n\text{-Pr})_2$	113.87	0.408			0.624	0.413
37	Et_2CMe_2	95.53	0.332			0.504	0.331
38	iso-$PrCMe_3$	91.61	0.313			0.472	0.309
39	*tert*-BuO(iso-Pr)	99.89	0.351			0.536	0.366
40	$EtMe_2C$(iso-Pr)	103.14	0.365			0.542	0.357

TABLE IVa. $CH_2R_AR_B$ Series and Their $\Sigma\sigma_{s^\circ}$(mono), σ_{s°(cal) and S°_{298}(g)/e.u. (cal)

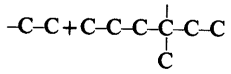
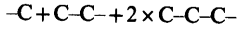
R_A	R_B	$\Sigma\sigma_{s^\circ}$(mono)	σ_{s°(cal)	S°_{298}(g)(cal)
Ph	COOH	0.417	0.342	97.85
Me	COOH	0.272	0.234	76.31
Me	$CONH_2$	0.282	0.242	77.65
Ph	Ph	0.472	0.383	107.54
Ph	CH_2Ph	0.523	0.421	117.37
OPh	Ph	0.523	0.421	117.37
NH_2	COOH	0.296	0.252	79.53

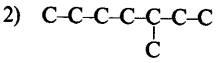
TABLE IVb. $R_ACH_2CH_2R_B$ Series and Their $\Sigma\sigma_{s^\circ}$(mono), σ_{s°(cal) and S°_{298}(g)/e.u. (cal)

R_A	R_B	$\Sigma\sigma_{s^\circ}$(mono)	σ_{s°(cal)	S°_{298}(g)(cal)	S°_{298}(g)(obs)[3,4]
Ph	COOH	0.469	0.381	107.04	
Me	COOH	0.353	0.295	87.71	
Me	$CONH_2$	0.343	0.287	86.21	
Me	F	0.242	0.212	72.54	72.71
Me	Cl	0.262	0.227	75.09	75.27
Me	Br	0.280	0.240	77.37	79.08
Me	I	0.294	0.251	79.35	80.32
Ph	CH_2Ph	0.572	0.458	127.72	

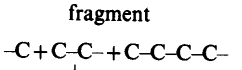
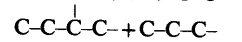
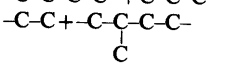
1) 

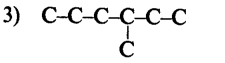
S°_{298} (g) = 124.47 e.u. $\sigma_{s^\circ} = 0.447$

reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-3		0.161 + 0.371	= 0.532	0.428	119.37
C-4		0.091 + 0.161 + 2 × 0.221	= 0.694	0.462	128.90
			average	0.445	124.14

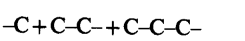
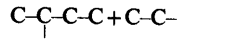
2) 

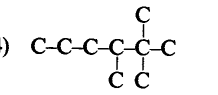
S°_{298} (g) = 110.32 e.u. $\sigma_{s^\circ} = 0.394$

reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-3		0.091 + 0.161 + 0.273	= 0.525	0.373	105.20
C-4		0.266 + 0.221	= 0.487	0.395	110.50
C-5		0.161 + 0.309	= 0.470	0.382	107.33
			average	0.383	107.68
when the correction of racemate is added, we get		$R \ln 2 = 1.38$ e.u.		0.389	109.05

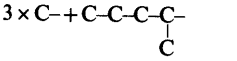
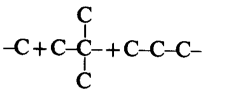
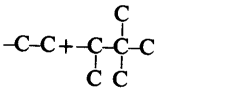
3) 

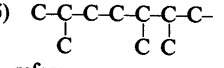
S°_{298} (g) = 101.37 e.u. $\sigma_{s^\circ} = 0.357$

reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-3		0.091 + 0.161 + 0.221	= 0.473	0.339	97.23
C-4		0.161 + 0.266	= 0.427	0.350	99.69
			average	0.345	98.46
when the correction of racemate is added, we get		$R \ln 2 = 1.38$ e.u.		0.351	99.84

4) 

S°_{298} (g) = 111.34 e.u. $\sigma_{s^\circ} = 0.398$

reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-2		0.273 + 0.310	= 0.583	0.386	108.22
C-3		0.091 + 0.210 + 0.221	= 0.522	0.371	104.72
C-4		0.161 + 0.313	= 0.474	0.385	108.07
			average	0.381	107.00
when the correction of racemate is added, we get		$R \ln 2 = 1.38$ e.u.		0.386	108.38

5) 

S°_{298} (g) = 122.90 e.u. $\sigma_{s^\circ} = 0.441$

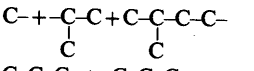
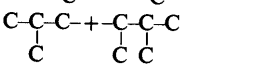
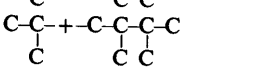
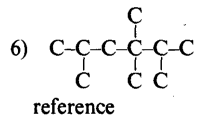
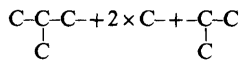
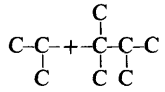
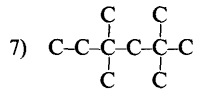
reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-3		0.091 + 0.199 + 0.310	= 0.600	0.423	117.86
C-4		0.266 + 0.293	= 0.559	0.448	125.03
C-5		0.199 + 0.347	= 0.546	0.439	122.27
			average	0.437	121.68
when the correction of racemate is added, we get		$R \ln 2 = 1.38$ e.u.		0.442	123.06

Chart 1. (Continued)

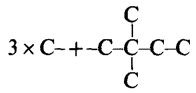
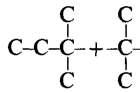
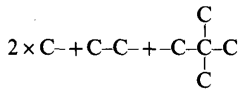


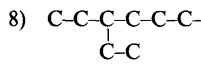
6) S°_{298} (g) = 118.25 e.u. $\sigma_{s^\circ} = 0.424$

reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-3		0.266 + 0.182 + 0.199	= 0.647	0.430	119.70
C-4		0.199 + 0.313	= 0.512	0.413	115.34
			average	0.422	117.52

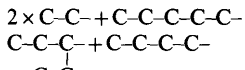
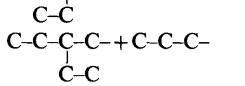
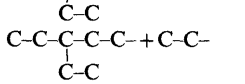


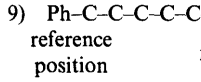
7) S°_{298} (g) = 115.91 e.u. $\sigma_{s^\circ} = 0.416$

reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-2		0.273 + 0.332	= 0.605	0.401	112.04
C-3		0.284 + 0.210	= 0.494	0.400	111.84
C-4		0.182 + 0.161 + 0.284	= 0.627	0.416	115.99
			average	0.406	113.29



8) S°_{298} (g) = 127.74 e.u. $\sigma_{s^\circ} = 0.458$

reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)		σ_{s° (cal)	S°_{298} (g) (cal)
C-3		0.322 + 0.319	= 0.641	0.450	125.41
C-4		0.309 + 0.273	= 0.582	0.466	130.06
C-5		0.344 + 0.221	= 0.565	0.453	126.32
C-6		0.391 + 0.161	= 0.552	0.443	123.54
			average	0.453	126.33



9) S°_{298} (g) = 151.71 e.u. $\delta_{s^\circ} = 0.532$

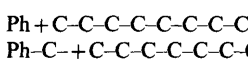
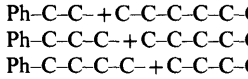
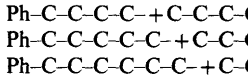
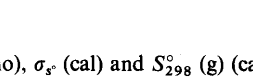
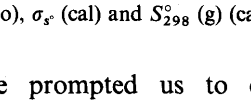
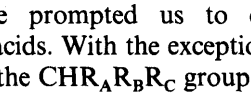
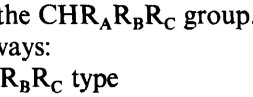
reference position	fragment	$\Sigma\sigma_{s^\circ}$ (mono)	σ_{s° (cal)	S°_{298} (g) (cal)
C-1		0.236 + 0.434 = 0.670	0.531	151.25
C-2		0.287 + 0.399 = 0.686	0.543	155.46
C-3		0.333 + 0.361 = 0.694	0.549	157.61
C-4		0.373 + 0.319 = 0.692	0.548	157.07
C-5		0.410 + 0.273 = 0.683	0.541	154.66
C-6		0.444 + 0.221 = 0.665	0.527	149.95
C-7		0.476 + 0.161 = 0.637	0.507	142.93
		average	0.535	152.70

Chart 1. Optimal $\Sigma\sigma_{s^\circ}$ (mono), σ_{s° (cal) and S°_{298} (g) (cal)

hitherto obtained have prompted us to estimate the descriptor σ_{s° of amino acids. With the exception of glycine, they are all classified in the $CHR_AR_BR_C$ group. Estimations are carried out by two ways:

1) substrate as $CHR_AR_BR_C$ type

2) substrate as $CH_2R_AR_B$ type

where $\sigma_{s^\circ}[-CH(NH_2)COOH] = 0.285$ is employed. The results are all summarized in Table V, where the descriptor $\sigma_{s^\circ}(R)$ given by the substituent R on *tert*-carbon becomes linear with σ_{s°(cal) inherent to the substrate.

For instance, lysine H_2N–CH_2–CH_2–CH_2–CH_2–CH-(NH_2)COOH gives descriptor σ_{s° by two ways 1) and 2) as mentioned below;

1) estimation of σ_{s° for *n*-pentyl amine classified to $CH_2R_AR_B$ is carried out under the conditions given below;

	fragment	$\Sigma\sigma_{s^\circ}$(mono)	σ_{s°(cal)
a.	N–C + C–C–C	0.184 + 0.221 = 0.405	0.334
b.	N–C–C + C–C	0.241 + 0.161 = 0.402	0.331

Mathematical average gives σ_{s°(C–C–C–C–N) = 0.333.

Then, lysine affords $\Sigma\sigma_{s^\circ}$(mono) = 0.333 + 0.115 + 0.181 = 0.629, and, using Eq. 2, we get the descriptor σ_{s°(cal) = 0.442.

2) σ_{s°(cal) = 0.285 for alanine means $\sigma_{s^\circ}[CH(NH_2)$-COOH] = 0.285. Using Eq. 2, we get

TABLE V. Amino Acids R–CH(NH_2)COOH and Their QSAR Descriptors

	R	σ_{s°(R)	$\Sigma\sigma_{s^\circ}$(mono)	σ_{s°(cal)	π[7]	I
Gly	H	0	0.296	0.252	0	0
Ala	Me	0.091	0.387	0.282	0.31	0
Val	iso-Pr	0.199	0.495	0.354	1.22	0
Leu	iso-Bu	0.266	0.562	0.398	1.70	0
Ile	*sec*-Bu	0.266	0.562	0.398	1.80	0
Ser	$HOCH_2$–	0.182	0.478	0.343	−0.04	1
Thr	MeCH(OH)–	0.221	0.517	0.369	0.26	1
Asp	$HOOC–CH_2$–	0.235	0.531	0.378	−0.77	2
Asn	$H_2NOC–CH_2$–	0.235	0.531	0.378	−0.60	2
Glu	$HOOC–CH_2CH_2$–	0.285	0.591	0.417	−0.64	2
Gln	$H_2NOC–CH_2CH_2$–	0.285	0.591	0.417	−0.22	2
Cys	$HS–CH_2$–	0.201	0.497	0.355	1.54	1
Met	$Me–S–CH_2CH_2$–	0.300	0.596	0.421	1.23	1
Lys	$H_2N–(CH_2)_4$–	0.330	0.626	0.445	−0.99	3
Hyl	$H_2N–CH_2CHCH_2CH_2$– (OH)	0.365	0.666	0.467		
Phe	$Ph–CH_2$–	0.287	0.583	0.412	0.79	1
Tyr	$HO–C_6H_4–CH_2$–	0.315	0.611	0.430	0.96	1
tert-Leu	*tert*-Bu	0.210	0.506	0.361		
γ-Me-Leu	Neopentyl	0.284	0.580	0.410		
nor-Val	*n*-Pr	0.221	0.517	0.370		
nor-Leu	*n*-Bu	0.273	0.569	0.403		
homo-Ala	Et	0.161	0.457	0.322		

homo-alanine R = Et $\Sigma\sigma_{s^\circ}$(mono) = 0.376 σ_{s°(cal) = 0.312

nor-valine R = *n*−Pr $\Sigma\sigma_{s^\circ}$(mono) = 0.446 σ_{s°(cal) = 0.364

The estimation is carried out under the conditions given below;

	fragment	$\Sigma\sigma_{s^\circ}$(mono)	σ_{s°(cal)
a.	N–C–C–C +0.285	0.286 + 0.285 = 0.571	0.457
b.	N–C–C +0.312	0.241 + 0.312 = 0.553	0.444
c.	N–C +0.364	0.184 + 0.364 = 0.548	0.440

The mathematical average gives $\sigma_{s^\circ} = 0.447$.

Finally, the descriptor $\sigma_{s^\circ} = 0.445$ is estimated for lysine.

Furthermore, the π-values of amino acids determined by Fauchére and Pliska[7] were found to be expressed by the linear combination of the independent variable σ_{s°(R) or σ_{s°(cal) and hydrogen bonding index ($I = 0, 1, 2$) as shown by Eq. 4.

$$\pi = 11.656(6.382)\sigma_{s^\circ}(\text{cal}) - 1.124(0.342)I + 2.789(2.184) \quad (4)$$

$n = 16$, $r = 0.909$, $s = 0.442$, $F = 26.25$

This result suggests that the hydrophobic substituent constant π could be expressed by the sum of contributions from the dispersion, repulsion, and averaged hydration number.

Conclusion

A logical estimation of the QSAR descriptor σ_{s° for $CH_2R_AR_B$, $CHR_AR_BR_C$ and $CR_AR_BR_CR_D$ could be carried out by means of the regression analyses using $\Sigma\sigma_{s^\circ}$(mono) as an independent variable. The term $\Sigma\sigma_{s^\circ}$(mono), composed of the sum of fragments, should be taken as the maximum or mean value for the branched- or straight-chain substrate, and corrections in the presence of the plane of symmetry, *trans* and *gauche* isomers, and optical antipode should be taken into account.

This method could be extended for an aliphatic substrate having a substituted phenyl ring in the molecular framework, and successful evidence is observed in the evaluation of the π-value of amino acids.

References

1) Y. Sasaki, H. Kawaki, and T. Takagi, *Chem. Pharm. Bull.*, **39**, 349 (1991).
2) H. Kawaki, *Chem. Pharm. Bull.*, "accepted."
3) S. A. Kudchadker and A. P. Kudchadker, *J. Phys. Chem. Ref. Data*, **7**, 1285 (1978).
4) D. R. Stull, E. F. Westrum, Jr., and G. C. Sinke, "The Chemical Thermodynamics of Organic Compounds," Willey, New York, 1969.
5) T. Takagi, T. Shindo, N. Iwata, A. Iwata, K. Tange, S. Fujii, and Y. Sasaki, "Proceedings of the 4th Software Conference," Osaka, 1988, p. 285.
6) Y. Sasaki, T. Takagi, and H. Kawaki, *Chem. Pharm. Bull.*, **36**, 3743 (1988).
7) J.-L. Fauchére and V. Pliska, *Eur. J. Med. Chem. Ther.*, **18**, 369 (1983).