Extended Additivity Model of Parameter log(L^{16})

Pavel Havelec and Jiří G. K. Ševčík[a)]

Department of Analytical Chemistry, Charles University, Albertov 2030, CZ-128 40 Prague 2, The Czech Republic

Received March 19, 1996; revised July 4, 1996

An additivity model of the apolar solute-solvent parameter log(L^{16}) was verified using sets of 939 nonaromatic and 1075 aromatic compounds. Unbiased distributions of errors and of the contribution significance level were statistically tested. An analysis of the CH_2 group contribution in 34 homologous series indicates that the differences among the homologous series are statistically insignificant and related to interactional contributions rather than to the nature of the CH_2 group.

Key words: additivity model; apolar solute descriptor; contribution of structure and interactions; gas-liquid chromatography; gas-liquid partition coefficient in *n*-hexadecane; homologous series; multiparametrical linear regression.

Contents

List of Tables

List of Figures

1. Introduction

The solute-solvent interactions in gas chromatography have been discussed in many papers over the last thirty years.[1] The most recent works have been published by Abraham,[2–5] Carr and coworkers[9–11] and Poole and coworkers.[12,13] The authors assume linear solvation energy relationships; therefore, the principal solvation equation can be written in the form,

$$\log(SP) = l \times \log(L^{16}) + r \times R_2 + s\pi_2^H + a \times \alpha_2^H + b \times \beta_2^H + SP_0, \quad (1)$$

where the dependent variable SP can be any retention characteristic (e.g., gas-liquid partition coefficient K, specific retention volume V_g, adjusted retention time t'_R or volume V'_R). The independent variables log(L^{16}), R_2, π_2^H, α_2^H, and β_2^H are parameters representing the properties of the individual solutes.

The regression coefficients l, r, s, a, b, and SP_0 in Eq. (1) characterize the stationary phases, i.e., they specify the ability of a phase to interact with a solute in a certain way. Equation (1) is very similar to the equation recommended by

a)Electronic mail: sevcik@prfdec.natur.cuni.cz

Carr's group[9] except for the product $r \times R_2$. Carr and coworkers[9] use the Taft–Kamlet[14] original product $d \times \delta_2$, where the solute parameter δ_2 is an empirical polarizability correction factor. This one is simpler as R_2, but it does not provide detailed differentiation in solute polarizability.[5]

The product $l \times \log(L^{16})$, where the solute descriptor $\log(L^{16})$ is the decadic logarithm of the partition coefficient for transfer of the solute from the gas phase to *n*-hexadecane at 298 K, is a combination of a cavity term with a general dispersion interaction term and is most decisive for the overall solute-solvent behavior.[8] The regression coefficient l reflects the ability of the stationary phase to separate adjacent members of a homologous series and exhibits a reasonable correlation with the partial Gibbs free energy of solvation of a methylene group into a stationary phase.[4]

The parameter $\log(L^{16})$ can be obtained experimentally, can be back-calculated from Eq. (1), and calculated using the additivity method.[1] Experimental methods using packed, capillary and head-space techniques are discussed in Carr's recent work,[8] concluding that:

(1) head-space gas chromatography leads to log L^{16} values free from interfacial adsorption,
(2) log L^{16} is a measure of the cavity formation process for a solute in *n*-hexadecane,
(3) log L^{16} includes contributions from both dispersive and dipolar interactions between the solute and *n*-hexadecane,
(4) the size of the solute and its functional group affect the extent of solvation of a solute in *n*-hexadecane,
(5) with *n*-hexadecane, the Gibbs free energies of CH_2 group transfer are not independent of the kind of homologous series,
(6) the gas-liquid Gibbs free energy of transfer of a functional group to *n*-hexadecane is related to its size and dipolarity-polarizability.

A comparison[7] of retention data corrected for the effect of interfacial adsorption shows that, for most solutes, these data are compatible and need not be corrected. An important result of this work is the finding that the results obtained for carboxylic acids are the extent of dimerization at 25 °C.

In our previous work,[1] an additivity model for prediction of the $\log(L^{16})$ parameter for nonaromatic compounds was derived:

$$\log(L^{16}(X)) = \sum_i l_i \times FG_i + \sum_j m_j \times SC_j + \sum_k n_k \times IC_k, \quad (2)$$

where

FG, *SC*, *IC* are the number of particular groups (*FG*) forming compound *X*, structural contributions (*SC*), interactional contributions (*IC*), respectively,

i, *j*, *k* are the identification number of group (*i*), of a structural contribution (*j*), of an interactional contribution (*k*), respectively.

l, *m*, *n* are the regression coefficients for the contributions of particular groups (*l*), structural contributions (*m*), interactional contributions (*n*), respectively.

In this work, the additivity model according to Eq. (2) has been extended to additional functional groups and aromatic compounds and applied to a large data set of 939 nonaromatic and 1075 aromatic compounds. Further, Carr's hypothesis of a selective CH_2 group contribution to the $\log(L^{16})$ value[8] was examined.

2. Procedures

Four data sets have been subjected to multilinear regression analysis (MLRA); set A consists of 1075 aromatic compounds, set N_f of 939 nonaromatic compounds, set N_t of 208 nonaromatic compounds and set N_c of 731 nonaromatic compounds. The data sets have been found in the literature[2] and/or kindly provided by Abraham.[6]

The investigated chemical compounds were redefined in terms of the number of specified functional groups, interactional contributions, and structural contributions representing the particular compound, according to Eq. (2). The contributions are listed in Tables 2 to 7. For the aromatic compounds it was assumed that the substituent atoms not directly bound to the aromatic ring were not influenced by the aromatic ring itself and the contributions of such functional groups were considered to be the same as in nonaromatic compounds.

The MLRA method was applied to the analysis of the matrix formed by the components of the particular data set and the specified contributions. Statistically insignificant results were omitted from the contributional matrix and the MLRA analysis was repeated.

Statistical evaluation was performed using the *S*-Plus program and the following tests:

(1) F-test for regression equation fit significance,
(2) t-test for significance of the difference between the regression coefficients from different regression equations.

TABLE 1. Statistical parameters of MLRA for the test data sets

Parameters	Data set			
	N_f	N_t	N_c	A
No. of compounds	939	208	731	1075
Degrees of freedom	882	151	674	1001
$\log(L^{16})_{min}$	−0.800	0.226	−0.800	1.830
$\log(L^{16})_{max}$	29.973	22.383	29.973	14.712
$\log(L^{16})_{mean}$	3.924	3.729	3.979	6.195
No. of functional groups	39	39	39	36
No. of structural contributions	3	3	3	8
No. of interactional contributions	15	15	15	30
Correlation coefficient (r)	0.999	0.999	0.998	0.995
Standard deviation (s)	0.122	0.102	0.140	0.208
Maximum error (me)	0.953	0.557	1.101	1.247
Mean statistical error (v%)	3.105	2.739	3.526	3.356
F-test (F)	6114.0	1022.9	3835.1	1333.8

TABLE 2. List of functional groups, their values, standard deviations (S.D.), fragment frequency (F.Freq.) and compound frequency (C.Freq.) for nonaromatic compounds (set N_f)

Name	Remarks	Index	Value	S.D.	F.Freq.	C.Freq.
$-CH_3$		11	0.340	0.006	1688	761
$-CH_2-$		12	0.502	0.001	2870	635
c $-CH_2-$	part of NAR[a]	13	0.532	0.003	772	162
-CH<		14	0.467	0.012	345	271
c -CH<	part of NAR	15	0.416	0.016	158	98
>C< , c >C<		16	0.443	0.021	191	133
$H_2C=$		17	0.249	0.016	95	87
-HC=		18	0.504	0.013	125	95
c -HC=	part of NAR	19	0.486	0.008	97	40
>C=		110	0.658	0.020	39	33
c >C=	part of NAR	111	0.573	0.028	24	22
HC≡		112	0.090	0.042	14	13
-C≡		113	0.592	0.038	16	14
-F		114	0.018	0.018	207	40
-Cl		115	0.799	0.015	144	77
c -Cl	connected to NAR	116	1.015	0.058	5	5
-Br		117	1.251	0.025	43	31
c -Br	connected to NAR	118	1.380	0.056	6	6
-I		119	1.752	0.038	11	11
-O-		120	0.285	0.016	104	73
c -O-	part of NAR	121	0.425	0.025	27	21
-CHO		122	0.984	0.025	27	27
-CO-		123	1.071	0.021	46	46
c -CO-	part of NAR	124	1.292	0.036	16	16
HCOO- , -COO-		125	1.071	0.013	159	154
-CN		126	1.163	0.031	15	14
$-NH_2$		127	0.859	0.028	19	18
>NH		128	0.590	0.034	16	15
>N-		129	0.347	0.045	9	9
$-NO_2$		130	1.535	0.042	9	9
HCON< , -CON<		131	2.363	0.047	8	8
-CONH-		132	2.436	0.041	13	13
-COOH		133	1.592	0.033	15	15
p -OH	primary alcohol	134	0.752	0.016	63	60
s -OH	secondary alcohol	135	0.710	0.021	43	42
t -OH	tertiary alcohol	136	0.585	0.034	17	17
-SH , -S-		137	1.298	0.020	48	45
c -S-	part of NAR	138	1.591	0.089	2	2
PO_4		139	2.027	0.090	2	2

[a]NAR—nonaromatic ring.

3. Results and Discussion

The results are given in Tables 1 to 10.

The original data set (split into four subsets N_f, N_t, N_c and A) has the following characteristics:

N_f—(full) set of all nonaromatic compounds ($N_f = N_t + N_c$),

N_t—(training) set of selected nonaromatic compounds containing all group, structural contribution and interactional contribution types like the set N_f,

N_c—(confirmation) set of nonaromatic compounds not included in set N_t,

A—set of all aromatic compounds.

TABLE 3. List of structural contributions, their values, standard deviations (S.D.), fragment frequency (F.Freq.) and compound frequency (C.Freq.) for nonaromatic compounds (set N_f)

Name	Remarks	Index	Value	S.D.	F.Freq.	C.Freq.
5-Ring	5-membered NAR[a]	m1	-0.067	0.021	56	46
6-Ring	6-membered NAR	m2	-0.022	0.020	103	91
cis X^1-CH=CH-X^2	*cis*-isomer	m3	0.112	0.040	12	12

[a]NAR—nonaromatic ring.

TABLE 4. List of interactional contributions, their values, standard deviations (S.D.), fragment frequency (F.Freq.) and compound frequency (C.Freq.) for nonaromatic compounds (set N_f)

Name	Remarks	Index	Value	S.D.	F.Freq.	C.Freq.
$-CX^1R^1-CX^2R^2-$	two alkyl groups on two adjacent C atoms	n1	0.119	0.013	62	38
$>CF_2$	two F atoms on one C atom	n2	-0.196	0.020	142	29
$>CCl_2$	two Cl atoms on one C atom	n3	-0.160	0.018	66	30
>C=CX-Cl	Cl atom on double bond C atom	n4	-0.149	0.024	15	7
$>CBr_2$	two Br atoms on one C atom	n5	-0.147	0.026	14	5
>C=CX-Br	Br atom on double bond C atom	n6	-0.186	0.048	6	4
>CFCl	F and Cl atoms on one C atom	n7	-0.155	0.016	29	11
>CFBr	F and Br atoms on one C atom	n8	-0.199	0.037	6	3
>CClBr	Cl and Br atoms on one C atom	n9	-0.130	0.032	10	6
-O >C< -O	two ether O atoms on one C atom	n10	-0.110	0.029	33	28
CH_3-O-	methoxy-bond	n11	0.164	0.021	37	27
-COO-C_nH_{2n}-OOC- (n=1-3)	alkylene dicarboxylate connection	n12	0.151	0.067	4	4
-NH-CH_3	sec. aminomethyl-bond	n13	0.195	0.054	5	4
>COH-COH<	hydroxyl group on two adjacent C atoms (1,2-diols)	n14	0.208	0.061	5	5
-S-CH_3	methylthio-bond	n15	0.151	0.040	8	6

The aim of splitting the original data set N_f into training N_t and confirming sets N_c was to prove the hypothesis of an unbiased selection of the input data.

First, MLRA was carried out independently for sets N_f and N_t. Based on the results for the N_t set, the $\log(L^{16})$ values were predicted for set N_c. The full and training sets were statistically compared using the t-test, according to Eq. (3):

$$t_i = \frac{X_i^{N_f} - X_i^{N_t}}{(s(X_i^{N_f})^2 + s(X_i^{N_t})^2)^{0.5}}, \tag{3}$$

where $X_i^{N_f}$, $X_i^{N_t}$ are the regression coefficients and $s(X_i^{N_f})^2$, $s(X_i^{N_t})^2$ their variances. It can be seen in Table 8 that both the values of the regression coefficients and their standard deviations are similar and the resulting t-test and probability levels demonstrate the similarity of sets N_f and N_t and thus an unbiased input data selection.

The decision to use a certain type of contribution (functional, structural, and interactional) in a particular regression model was determined by the following criteria:

(1) particular contribution had to have a compound frequency ≥ 2,
(2) the uncertainty that the magnitude of the calculated contribution did not significantly differ from zero should be:
  (a) greater than 5% for functional contributions,
  (b) greater than 20% for structural and interactional contributions,
(3) the uncertainty that the magnitude of the calculated contribution was the same as for a similar next contribution should be less than 5%.

Exceptions from the above rules were tolerated only for the functional contribution of fluorine -F($P_{(-F=0)}$ = 30.8%), structural and interactional contribution of the aliphatic 6-ring ($P_{(6\text{-ring}=0)}$ = 27.3%) and exchange of a primary and secondary OH group (primary OH≈secondary OH).

The hypothesis of the equality of predicted and experimental values [$\log(L^{16})_{calc} = \log(L^{16})_{meas}$] was tested by F-test according to Eq. (4):

$$F = \frac{r^2 \times (n-p)}{(1-r^2) \times (p-1)}, \tag{4}$$

where r is the correlation coefficient, n is the number of compounds used in regression, p is the number of regression coefficients. The results of statistical evaluation for all four sets are shown in Table 1.

It can be seen in Table 1 that all tested dependencies are statistically significant. Somewhat poorer results for data set A indicate that most probably there can exist some additional structural and interactional contributions in the model, especially ortho-interactions, that have not yet been evaluated because of an insufficient compound frequency. The plots of the measured versus the calculated $\log(L^{16})$ values for set N_f and set A are shown in Figs. 1 and 3. The histograms of estimation errors for sets N_f and A are shown in Figs. 2 and 4, where near-Gaussian error distribution curves centered at zero are seen.

Great attention was paid to the distribution of the differences between the measured and calculated values. It has been found that the differences are normally distributed with the mean value equal to zero. Thus it can be concluded that the estimated values are not biased and that the additivity concept is correct.

Carr *et al.*[8] have proposed a hypothesis of a selective methylene group contribution as a function of the type of the homologous series (conclusion No. 5[8]). We selected 29 homologous series with at least 5 compounds from the set N

TABLE 5. List of functional groups, their values, standard deviations (S.D.), fragment frequency (F.Freq.) and compound frequency (C.Freq.) for aromatic compounds (set A)

Name	Remarks	Index	Value	S.D.	F.Freq.	C.Freq.
))CH	part of AR[a]	140	0.498	0.013	5394	1063
))C-	part of AR	141	0.753	0.021	2308	1001
))C((	part of AR	142	0.753	0.024	948	247
))C((	part of AR	143	0.821	0.053	66	32
))N	part of AR	144	0.735	0.025	179	170
))NH	part of AR	145	1.825	0.110	29	29
))N-	part of AR	146	1.475	0.118	6	6
))O	part of AR	147	0.381	0.080	11	11
))S	part of AR	148	1.442	0.097	35	34
a -CH_3	connected to AR	149	0.228	0.020	645	420
a -CH_2-	connected to AR	150	0.307	0.021	257	213
ac -CH_2-	NAR part connected to AR	151	0.432	0.025	108	58
a -CH<	connected to AR	152	0.190	0.037	50	47
ac -CH<	NAR[b] part connected to AR	153	0.206	0.069	15	15
a >C<	connected to AR	154	0.251	0.054	22	21
a -HC=	connected to AR	155	0.518	0.041	39	36
ac -HC=	NAR part connected to AR	156	0.352	0.083	5	4
a >C= , ac>C=	connected to AR, NAR part connected to AR	157	0.456	0.127	3	3
a -F	connected to AR	158	-0.291	0.027	47	30
a -Cl	connected to AR	159	0.533	0.021	327	146
a -Br	connected to AR	160	0.908	0.037	36	29
a -I	connected to AR	161	1.399	0.046	20	17
a -O-	connected to AR	162	0.382	0.029	92	84
ac -O-	NAR part connected to AR	163	0.279	0.092	4	3
a -CHO	connected to AR	164	1.039	0.050	24	22
a -CO-	connected to AR	165	1.212	0.070	10	10
ac -CO-	NAR part connected to AR	166	1.130	0.058	11	7
a HCOO-	connected to AR	167	0.789	0.155	2	2
a -COO-	connected to AR	168	1.182	0.032	57	52
a -CN	connected to AR	169	0.923	0.057	16	16
a -NH_2	connected to AR	170	0.896	0.033	61	58
a >NH	connected to AR	171	0.749	0.050	20	19
a >N-	connected to AR	172	0.848	0.064	15	15
a -NO_2	connected to AR	173	1.314	0.034	66	58
a -OH	connected to AR	174	0.800	0.027	163	160
a -SH , a -S-	connected to AR	175	1.032	0.100	5	5

[a]AR—aromatic ring.
[b]NAR—nonaromatic ring.

TABLE 6. List of structural contributions, their values, standard deviations (S.D.), fragment frequency (F.Freq.) and compound frequency (C.Freq.) for aromatic compounds (set A)

Name	Remarks	Index	Value	S.D.	F.Freq.	C.Freq.
5-AR[a]	5-membered AR	m4	-0.479	0.083	95	95
6-AR	6-membered AR	m5	-0.127	0.068	1577	1049
7-AR	7-membered AR	m6	0.777	0.136	5	5
))C-C((	bond between two C atoms of adjacent ARs conjuct edge	m7	0.074	0.024	545	232
))C-R`-C((	connection of two ARs by one methylene group	m8	-0.185	0.057	11	9
))C-C((	bond between two independent ARs (biphenyl connection)	m9	-0.075	0.043	63	50
-NH-N=	two adjacent N atoms in one AR (structural part of pyrazole)	m10	-0.353	0.155	3	3
AR-CX^1=CX^2-CN	derivates of trans-cinnamonitrile	m11	0.307	0.081	9	9

[a]AR—aromatic ring.

TABLE 7. List of interactional contributions, their values, standard deviations (S.D.), fragment frequency (F.Freq.) and compound frequency (C.Freq.) for aromatic compounds (set A)

Name	Remarks	Index	Value	S.D.	F.Freq.	C.Freq.
$-CH_3$, $-CH_3$	ortho effect	n16	0.123	0.021	85	56
$-CH_3$, $-CH_2-$	ortho effect	n17	0.076	0.038	35	32
$-CH_3$, c $-CH_2-$	ortho effect	n18	0.150	0.102	5	5
$-CH_3$,)⟩CH	ortho effect	n19	0.111	0.026	59	47
$-CH_3$, AR [a]	ortho effect	n20	-0.516	0.060	11	9
AR , AR	ortho effect	n21	-0.890	0.105	3	2
AR ,)⟩C-	ortho effect	n22	-0.463	0.063	10	8
$-CH_3$, -Cl	ortho effect	n23	0.060	0.036	30	20
$-CH_3$, -O-	ortho effect	n24	-0.117	0.079	6	5
-Cl , -O-	ortho effect	n25	-0.099	0.041	32	23
-O- , -O-	ortho effect	n26	0.730	0.078	11	11
-Cl , -CHO	ortho effect	n27	0.214	0.065	10	7
-COO- , -COO-	ortho effect	n28	-0.495	0.163	2	2
-O- , $-NH_2$	ortho effect	n29	-0.142	0.106	5	5
$-CH_3$, -N<	ortho effect	n30	-0.408	0.098	4	3
-Cl , $-NO_2$	ortho effect	n31	-0.393	0.047	20	14
$-CH_3$, -OH	ortho effect	n32	-0.057	0.038	34	28
$-CH_2-$, -OH	ortho effect	n33	-0.268	0.051	17	14
-CH< , -OH	ortho effect	n34	-0.355	0.074	9	8
>C< , -OH	ortho effect	n35	-0.479	0.104	4	3
-F , -OH	ortho effect	n36	-0.148	0.099	3	2
-Cl , -OH	ortho effect	n37	-0.390	0.041	31	22
-Br , -OH	ortho effect	n38	-0.465	0.097	4	3
$-NO_2$, -OH	ortho effect	n39	-0.655	0.082	7	6
)⟩N , $-CH_3$	ortho effect	n40	-0.138	0.031	61	51
)⟩N , $-CH_2-$	ortho effect	n41	-0.203	0.040	34	29
)⟩N , c $-CH_2-$	ortho effect	n42	-0.412	0.130	3	3
)⟩N ,)⟩C	ortho effect	n43	-0.135	0.034	43	33
)⟩NH ,)⟩C	ortho effect	n44	-0.135	0.062	36	25
)⟩S ,)⟩C	ortho effect	n45	-0.230	0.065	30	26

[a]AR—aromatic ring.

and 5 homologous series with at least 4 compounds from the set A. Homologous series included compounds in which the CH_2 group was connected to another CH_2 group (e.g. *n*-pentane was the first evaluated homologue). We assumed that, in compounds not fulfilling the above condition, the contribution of the CH_2 group can be affected by interactional contributions. We tested the $\log(L^{16})$ dependence on the number of CH_2 groups n_c according to Eq. (5):

$$\log(L^{16}(X_i)) = a \times n_c(X_i) + b. \quad (5)$$

The results are given in Table 9. The values of the confidence interval of the slope at a 95% probability level are plotted in Fig. 5 for all the test homologous series. It can be seen that the mean value of the slope a calculated for all the test homologous series (0.500) is the same as that calculated from the set N_f (0.502). Figure 5 shows that only the value for alkanoic acids is out of order (a=0.543, a_{mean} = 0.500), most probably due to dimerization.[7] The tests of slope homogeneity were performed and their results were as follows:

(a) for the full set of 34 homologous series, the F-test shows that there cannot be applied common value of the mean slope at 95% probability significance level,

(b) for the set of 30 homologous series (without alkanoic acids, methyl alkanoates, ethyl alkanoates, ket-2-ones), the F-test shows that there is the mean slope value at 95% probability significance level.

A comparison of the slope a for Carr's[8] and our results for the same homologous series is shown in Table 10. It can be seen that the results are similar (CH_2(Carr)=0.498; CH_2(this work)=0.500), small differences being due to Carr's use of the short homologous series and the way of n_c carbon number determination (some more variance of the mean of our results is due to the larger number of the homologous series used and an extended range of n_c). Based on our additive model, $\log(L^{16})$ was calculated for the set of 88 compounds used by Carr (a comparison of the head-space data and the Abraham capillary column data). A correlation coefficient of 0.993 and F=6809 has been found demonstrating a good agreement between the most critically evaluated experimental data and the predicted data.

4. Conclusions

The model of the additivity of particular contributions to the $\log(L^{16})$ value has been verified and extended to a general description of nonaromatic and aromatic compounds. The parameter $\log(L^{16})$ can be estimated for any compound

TABLE 8. Test of statistical significance of differences in regression coefficients l_i, m_j, and n_k for full set N_f and training set N_t of nonaromatic compounds

Name	Index	Training set N_t		Full set N_f		t-test	
		Value	Standard error	Value	Standard error	$t_{exp.}$	Significance level (%)
$-CH_3$	l1	0.339	0.015	0.340	0.006	0.066	94.750
$-CH_2-$	l2	0.503	0.002	0.502	0.001	0.273	78.515
c $-CH_2-$	l3	0.530	0.006	0.532	0.003	0.250	80.300
-CH<	l4	0.459	0.027	0.467	0.012	0.270	78.741
c -CH<	l5	0.409	0.043	0.416	0.016	0.153	87.875
>C< , c >C<	l6	0.484	0.057	0.443	0.021	0.678	49.876
$H_2C=$	l7	0.256	0.039	0.249	0.016	0.173	86.294
-HC=	l8	0.492	0.025	0.504	0.013	0.408	68.419
c -HC=	l9	0.512	0.023	0.486	0.008	1.062	28.976
>C=	l10	0.664	0.054	0.658	0.020	0.103	91.813
c >C=	l11	0.435	0.086	0.573	0.028	1.524	12.949
HC≡	l12	0.095	0.091	0.090	0.042	0.054	95.733
-C≡	l13	0.580	0.061	0.592	0.038	0.157	87.574
-F	l14	-0.004	0.041	0.018	0.018	0.483	62.958
-Cl	l15	0.778	0.029	0.799	0.015	0.633	52.755
c -Cl	l16	0.961	0.125	1.015	0.058	0.398	69.099
-Br	l17	1.224	0.078	1.251	0.025	0.327	74.431
c -Br	l18	1.373	0.082	1.380	0.056	0.074	94.120
-I	l19	1.765	0.071	1.752	0.038	0.163	87.038
-O-	l20	0.293	0.035	0.285	0.016	0.211	83.297
c -O-	l21	0.404	0.043	0.425	0.025	0.422	67.384
-CHO	l22	0.997	0.052	0.984	0.025	0.224	82.337
-CO-	l23	1.035	0.042	1.071	0.021	0.763	44.654
c -CO-	l24	1.314	0.083	1.292	0.036	0.253	80.089
HCOO- , -COO-	l25	1.067	0.029	1.071	0.013	0.123	90.262
-CN	l26	1.069	0.049	1.163	0.031	1.612	10.895
$-NH_2$	l27	0.805	0.062	0.859	0.028	0.790	43.048
>NH	l28	0.698	0.075	0.590	0.034	1.314	19.077
>N-	l29	0.382	0.081	0.347	0.045	0.373	70.947
$-NO_2$	l30	1.501	0.086	1.535	0.042	0.365	71.550
HCON< , -CON<	l31	2.377	0.095	2.363	0.047	0.135	89.311
-CONH-	l32	2.452	0.086	2.436	0.041	0.167	86.774
-COOH	l33	1.559	0.070	1.592	0.033	0.433	66.536
p -OH	l34	0.758	0.035	0.752	0.016	0.163	87.090
s -OH	l35	0.764	0.047	0.710	0.021	1.054	29.346
t -OH	l36	0.572	0.066	0.585	0.034	0.179	85.822
-SH , -S-	l37	1.303	0.037	1.298	0.020	0.130	89.653
c -S-	l38	1.527	0.136	1.591	0.089	0.392	69.560
PO_4	l39	2.026	0.094	2.027	0.090	0.008	99.326
5-Ring	m1	0.012	0.063	-0.067	0.021	1.195	23.379
6-Ring	m2	-0.005	0.056	-0.022	0.020	0.281	77.881
cisX^1-CH=CH-X^2	m3	0.171	0.069	0.112	0.040	0.741	45.999
$-CX^1R^1-CX^2R^2-$	n1	0.149	0.036	0.119	0.013	0.769	44.314
$>CF_2$	n2	-0.195	0.050	-0.196	0.020	0.026	97.911
$>CCl_2$	n3	-0.160	0.047	-0.160	0.018	0.002	99.840
>C=CX-Cl	n4	-0.170	0.068	-0.149	0.024	0.297	76.720
$>CBr_2$	n5	-0.066	0.196	-0.147	0.026	0.409	68.295
>C=CX-Br	n6	-0.168	0.094	-0.186	0.048	0.172	86.344
>CFCl	n7	-0.118	0.028	-0.155	0.016	1.163	24.651
>CFBr	n8	-0.233	0.075	-0.199	0.037	0.404	68.650
>CClBr	n9	-0.094	0.074	-0.130	0.032	0.451	65.284
-O >C< -O	n10	-0.063	0.060	-0.110	0.029	0.697	48.714
CH_3-O-	n11	0.126	0.051	0.164	0.021	0.699	48.545
-COO-C_nH_{2n}-OOC- (n=1-3)	n12	0.098	0.097	0.151	0.067	0.448	65.486

TABLE 8. Test of statistical significance of differences in regression coefficients l_i, m_j, and n_k for full set N_f and training set N_t of nonaromatic compounds—Continued

Name	Index	Training set N_t		Full set N_f		t-test	
		Value	Standard error	Value	Standard error	$t_{exp.}$	Significance level (%)
-NH-CH_3	n13	0.165	0.140	0.195	0.054	0.201	84.096
>COH-COH<	n14	0.200	0.133	0.208	0.061	0.055	95.583
-S-CH_3	n15	0.173	0.093	0.151	0.040	0.213	83.129

with a high precision by adding the values for the functional groups, interactional and structural contributions in any compound.

The results demonstrate statistical significance of the magnitude of partial contributions and unbiased input data selection. Thus the computed contributions do not change with a data set extension. It has been shown that the calculated CH_2 group contribution is statistically the same as the mean value of the CH_2 group calculated from 34 homologous series, and it is assumed that the data scatter between particular homologous series is caused by interactional contributions and an effect of dipolarity parameter π_2^H. The use of the computed mean value should thus lead to a minimization of all effects other than dispersion ones.

This approach has a significant impact on the theoretical fields of expert systems and on the practical fields of medicine, pharmacy, and biochemistry, and permits an estimation

TABLE 9. Statistical parameters of Eq. (5) for the test homologous series

Name	Index	a	S.D.(a)	n	CH_2 Range	r	S.D.	F-test
n-Alkanes	1	0.506	0.0002	33	1-58	1.000	0.0176	5794543
2-Methylalkanes	2	0.488	0.0012	6	1-6	1.000	0.0050	163294
3-Methylalkanes	3	0.476	0.0036	5	1-5	1.000	0.0115	17229
2,2-Dimethylalkanes	4	0.469	0.0048	5	1-5	1.000	0.0152	9572
n-Alk-1-enes	5	0.496	0.0019	13	1-13	1.000	0.0262	65456
n-Alk-1-ynes	6	0.514	0.0053	8	1-9	1.000	0.0374	9425
1-Chloroalkanes	7	0.517	0.0040	6	1-6	1.000	0.0166	16972
1-Bromoalkanes	8	0.494	0.0062	7	1-7	1.000	0.0327	6417
Dialkylethers	9	0.488	0.0064	5	2-10	1.000	0.0403	5877
1-Alkanals	10	0.499	0.0025	12	1-12	1.000	0.0295	41060
Ket-2-ones	11	0.485	0.0019	10	1-16	1.000	0.0243	67614
Ket-3-ones	12	0.481	0.0047	5	2-6	1.000	0.0149	10378
Alkyl formates	13	0.531	0.0105	5	1-5	0.999	0.0331	2564
Alkyl acetates	14	0.506	0.0008	11	1-11	1.000	0.0085	391912
Alkyl butanoates	15	0.493	0.0033	5	1-5	1.000	0.0104	22558
Methyl alkanoates	16	0.483	0.0004	17	1-17	1.000	0.0084	1347797
Ethyl alkanoates	17	0.482	0.0002	17	1-17	1.000	0.0041	5739255
Alkyl isobutanoates	18	0.494	0.0036	5	1-5	1.000	0.0113	19037
Alkyl 2-methylbutanoates	19	0.482	0.0013	5	1-5	1.000	0.0041	137199
Alkanoic acids	20	0.543	0.0004	10	1-10	1.000	0.0033	2172680
1-Cyanoalkanes	21	0.481	0.0061	9	1-9	0.999	0.0469	6321
1-Nitroalkanes	22	0.505	0.0050	5	1-5	1.000	0.0157	10320
Alkylamines	23	0.486	0.0090	7	1-7	0.999	0.0475	2929
cis-N-4-alkylcyclohexylacetamides	24	0.525	0.0323	5	2-6	0.994	0.1021	265
trans-N-4-alkylcyclohexylacetamides	25	0.529	0.0330	5	2-6	0.994	0.1043	258
n-Alkan-1-ols	26	0.512	0.0027	11	1-11	1.000	0.0284	35692
n-Alkan-2-ols	27	0.500	0.0024	7	1-7	1.000	0.0124	45401
n-Alkan-1-thiols	28	0.520	0.0077	9	1-9	0.999	0.0596	4567
2-chloroethyl-alkylsulfides	29	0.515	0.0076	7	1-7	0.999	0.0403	4558
Alkylbenzenes	30	0.489	0.0034	9	1-10	1.000	0.0282	20685
Alkyl benzoates	31	0.515	0.0032	11	1-11	1.000	0.0332	26525
N-Alkylanilines	32	0.512	0.0001	4	1-7	1.000	0.0003	64615222
2(1-Methylalkyl)phenols	33	0.475	0.0096	4	2-5	1.000	0.0214	2465
4(1-Methylalkyl)phenols	34	0.490	0.0097	4	2-5	1.000	0.0217	2557
Average		0.500	0.0182	34	...	...	...	...

TABLE 10. Comparison of CH_2 contributions according to Carr[8] and our work

	Carr's work			Our work		
Homolog series	n_c range	a	S.D.(a)	n_c range	a	S.D.(a)
alkanes	5-10	0.499	0.003	1-58	0.506	0.0002
2-ketones	3-7	0.501	0.007	1-16	0.485	0.0019
aldehydes	3-8	0.503	0.011	1-12	0.499	0.0025
ethers	4-10	0.472	0.014	2-10	0.488	0.0064
formates	2-4	0.501	0.009	1-5	0.531	0.0105
acetates	3-8	0.485	0.006	1-11	0.506	0.0008
ethyl alkanoates	3-6	0.478	0.007	1-17	0.482	0.0002
branched acetates	5-7	0.532	0.014	···	···	···
alcohols	1-6	0.517	0.006	1-11	0.512	0.0027
nitroaliphatics	1-3	0.486	0.017	1-5	0.505	0.005
nitriles	1-5	0.500	0.004	1-9	0.481	0.006
average	···	0.498	0.009	···	0.500	0.0182

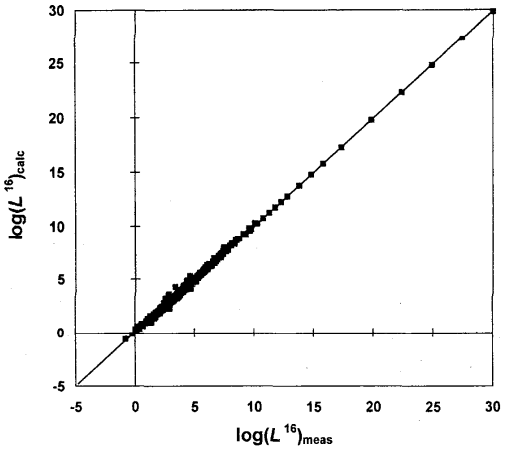

FIG. 1. Correlation between the measured and calculated $\log(L^{16})$ for dataset N_f.

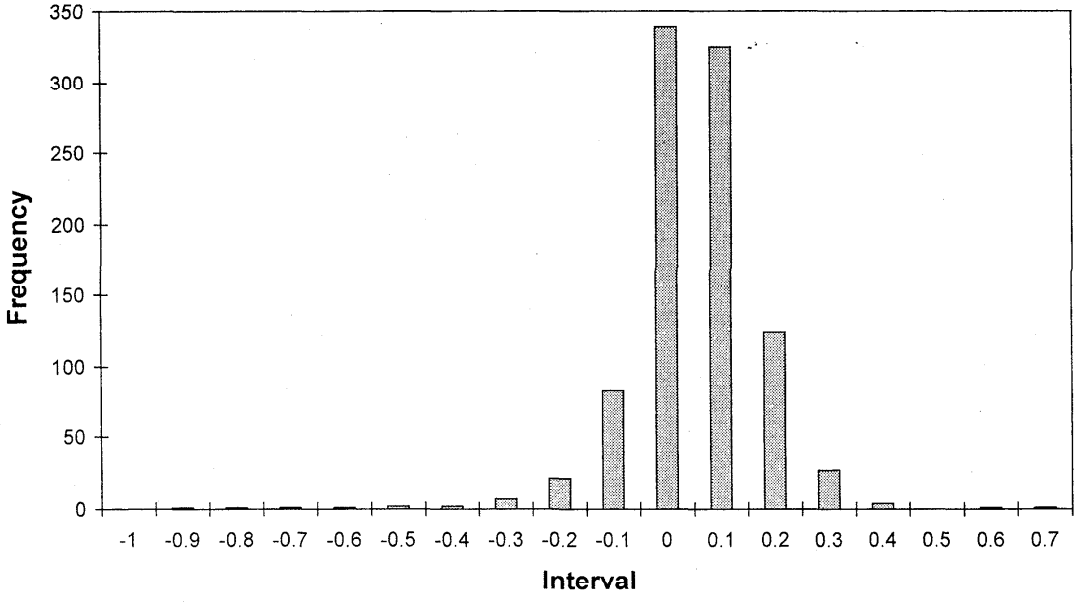

FIG. 2. Distribution of estimation error for dataset N_f.

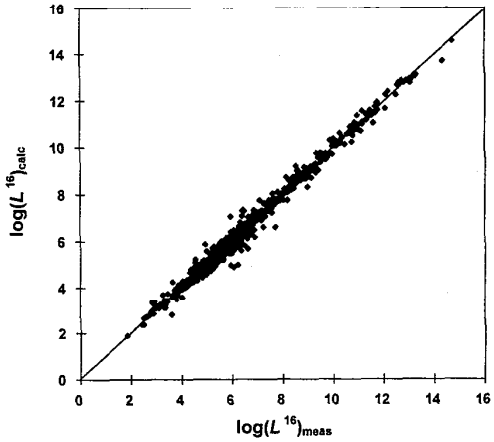

FIG. 3. Correlation between the measured and calculated $\log(L^{16})$ for dataset A.

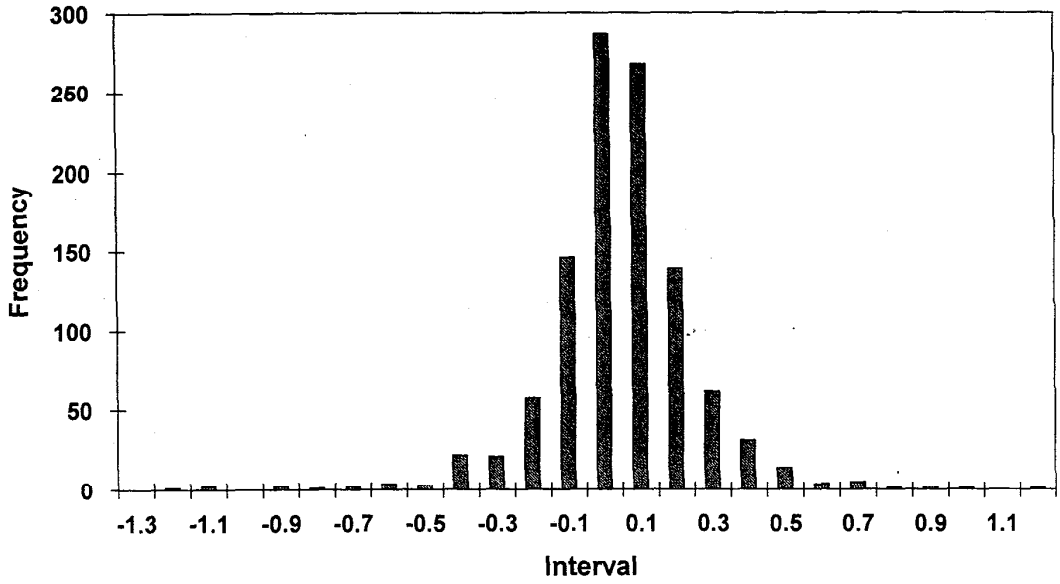

FIG. 4. Distribution of estimation error for dataset A.

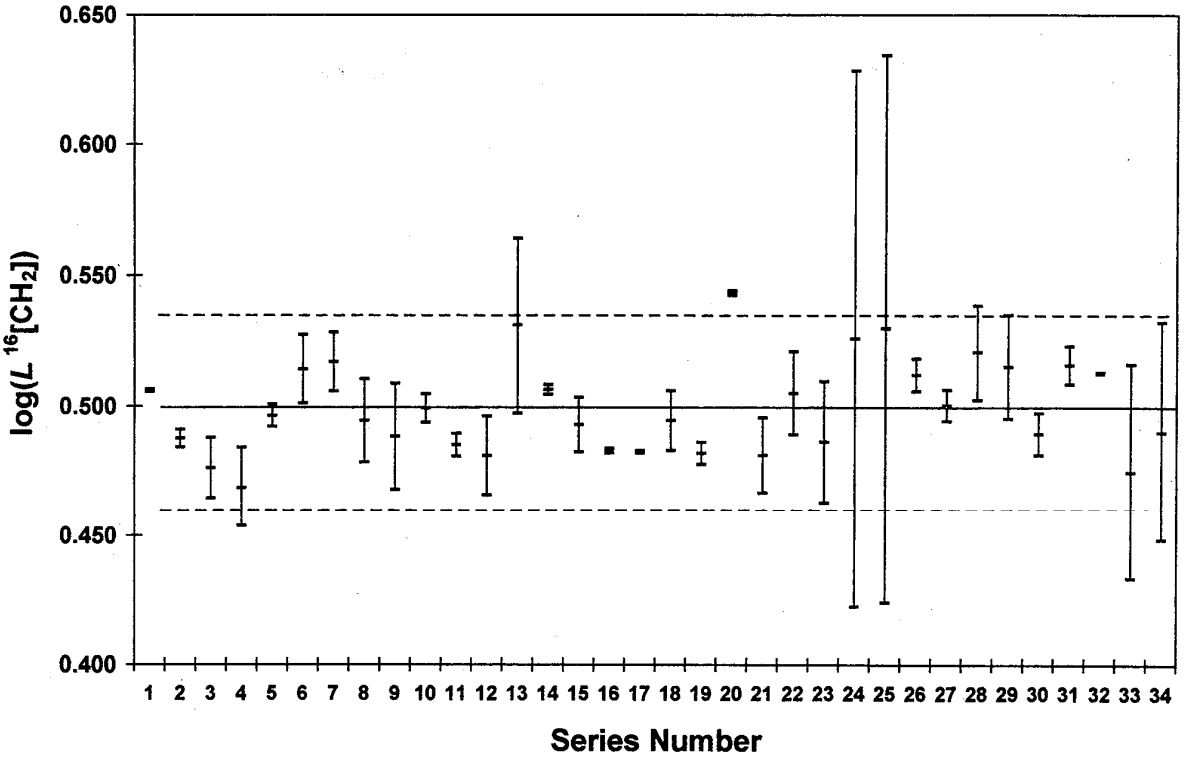

FIG. 5. Values of $\log(L^{16}[CH_2])$ for the test homologous series (see Table 9). The bar length corresponds to the 95% probability level. The mean value and its 95% confidence interval are shown in solid and dotted lines, respectively.

of the distribution of the effects of pharmaceuticals in physiological systems, i.e., blood-brain, etc.

5. Acknowledgments

The authors would like to thank Professor M.H. Abraham (Department of Chemistry, University College London, UK) for his kind provision of the original data set of the log (L^{16}) values. This work was supported by the grant from the Grant Agency of the Czech Republic, GACR No. 203/95/0805.

6. References

[1] P. Havelec and J. G. K. Ševčík, J. Chromatogr. **677**, 319 (1994).
[2] M. H. Abraham, J. Chromatogr. **644**, 95 (1993).
[3] M. H. Abraham, G. S. Whiting, R. M. Doherty, and W. J. Shuely, J. Chromatogr. **587**, 213 (1991).
[4] M. H. Abraham, G. S. Whiting, R. M. Doherty, and W. J. Shuely, J. Chromatogr. **587**, 229 (1991).
[5] M. H. Abraham, G. S. Whiting, R. M. Doherty, and W. J. Shuely, J. Chem. Soc. Perkin Trans. **2**, 1451 (1990).
[6] M. H. Abraham (private communications).
[7] M. H. Abraham, J. Andonian-Haftven, C. M. Du, J. P. Osei-Owusu, P. Sakellariou, W. J. Shuely, C. F. Poole, and S. K. Poole, J. Chromatogr. **688**, 125 (1994).
[8] A. J. Dallas and P. W. Carr, J. Phys. Chem. **98**, 4927 (1994).
[9] J. Li, Y. Zhang, A. J. Dallas, and P. W. Carr, J. Chromatogr. **550**, 101 (1991).
[10] J. Li, Y. Zhang, and P. W. Carr, Anal. Chem. **65**, 1969 (1993).
[11] J. Li, Y. Zhang, and P. W. Carr, Anal. Chem. **64**, 210 (1992).
[12] T. O. Kollie and C. F. Poole, J. Chromatogr. **556**, 457 (1991).
[13] T. O. Kollie, C. F. Poole, M. H. Abraham, and G. S. Whiting, Anal. Chim. Acta **259**, 1 (1992).
[14] R. W. Taft, J.-L. M. Abboud, and M. J. Kamlet, J. Am. Chem. Soc. **103**, 1080 (1981).