

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/231171114>

Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Qualitative Structure–Property Relationships Treatment

ARTICLE *in* ANALYTICAL CHEMISTRY · JUNE 1994

Impact Factor: 5.64 · DOI: 10.1021/ac00083a005

CITATIONS

127

READS

43

5 AUTHORS, INCLUDING:



[Victor Lobanov](#)

Covance

64 PUBLICATIONS 3,814 CITATIONS

SEE PROFILE



[Mati Karelson](#)

University of Tartu

187 PUBLICATIONS 7,374 CITATIONS

SEE PROFILE

Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Quantitative Structure–Property Relationship Treatment

Alan R. Katritzky,* Elena S. Ignatchenko, Richard A. Barcock, and Victor S. Lobanov

Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P.O. Box 117200, Gainesville, Florida 32611–7200

Matl Karelson

Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu EE 2400, Estonia

A general yet effective QSPR treatment on 152 individual structures incorporating a wide cross section of classes of organic compounds has given good six-parameter correlations for gas chromatographic retention times ($R^2_{cv} = 0.955$ for t_R) and for Dietz response factors ($R^2_{cv} = 0.881$ for RF_{Dietz}). The statistical treatment utilized the CODESSA program code to find the best multiparameter correlations from subsets of given size within larger sets of molecular descriptors. In the case of t_R , the most important descriptors are α polarizability and the minimum valency at an H atom, describing the dispersional and hydrogen-bonding interaction between the compound studied and the gas chromatographic medium, respectively. In the case of RF, the most important descriptors are the relative weight of "effective" carbon atoms and the total molecular one-center one-electron repulsion energy in the molecule. It has thus been demonstrated that with just six-parameter equations, in each case, the retention times and response factors for compounds that are unknown, unavailable, or not easily handled (toxic, odorous, etc.) can now be predicted with a considerable degree of confidence.

In extensive and ongoing investigations into the "aquathermolysis" of organic compounds at high temperatures,^{1,2} we have intensely utilized gas chromatography–mass spectroscopy (GC/MS) as the means to identify the components of and to quantitatively analyze complex reaction mixtures. The advantages of using GC/MS in these and other cases where many compounds are formed are particularly evident as, in principle, the method allows both the identification of the products and the determination of their proportions. However, quantitative estimations of compound proportions from the GC trace require knowledge of, or a method to estimate, the response factor for each compound under the GC experimental conditions employed. In our work, numerous products were unavailable as standards (some of them were unknown) and a method for estimating response factors was imperative. Although compound structure can usually be deduced from mass spectral (MS) fragmentation patterns, a method for accurately predicting retention times would also be of considerable assistance in preliminary identification of com-

ponents and would serve as useful supporting evidence for structure.

We now report the first comprehensive correlation of both retention times and response factors for a large and widely diversified set of organic compounds which has led to six-parameter equations which allow each of these quantities to be predicted for unknown compounds with considerable confidence.

Previous QSPR on GS Retention Indices. Although numerous sets of gas chromatographic retention indices have been successfully correlated by QSPR methods, without exception, each correlation relates to the compounds of a single class or of a few closely related classes. A brief summary of the most important literature work follows. An excellent overview of the theoretical basis of QSRR (quantitative structure retention relationships) to chromatography and of the molecular descriptors used therein has been given by Kaliszan,³ but most of his examples relate to the liquid phase.

Jurs et al. successfully demonstrated the computer-assisted prediction of the retention indices for diverse sets of (i) substituted pyrazines (widely known for their characteristic flavoring properties),^{4,5} (ii) polycyclic aromatic compounds (largest class of known chemical carcinogens),⁶ (iii) stimulants and narcotics,⁷ and (iv) anabolic steroids.⁸ In each case, the predictive models were generated through multiple regression analysis of calculated structural descriptors which encode topological, geometrical, electronic, and physical properties of the solute molecules in question. Examination of the models produced by the statistical analysis showed that the descriptors encode information related to the types of interactions that take place between the solute (and solvent) molecules on stationary phases during the separation processes. It was also shown that different descriptors became important in such predictive models as the polarity of the stationary phase is changed and, therefore, each phase should be modeled separately. The predictive ability of models generated in this

(3) Kaliszan, R. *Quantitative Structure-Chromatographic Retention Relationships*; Wiley-Interscience: New York, 1987.

(4) Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1989**, *61*, 1328.

(5) Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323.

(6) Whalen-Pedersen, E. K.; Jurs, P. C. *Anal. Chem.* **1981**, *53*, 2184.

(7) Georgakopoulos, C. G.; Kiburis, J. C.; Jurs, P. C. *Anal. Chem.* **1991**, *63*, 2021.

(8) Georgakopoulos, C. G.; Tsika, O. G.; Kiburis, J. C.; Jurs, P. C. *Anal. Chem.* **1991**, *63*, 2025.

(1) Siskin, M.; Katritzky, A. R. *Science* **1991**, *254*, 231.

(2) Katritzky, A. R.; Luxem, F. J.; Murugan, R.; Greenhill, J. V.; Siskin, M. *Energy Fuels* **1992**, *6*, 450.

fashion were shown to be accurate within a few percent of the mean retention index value, indicating that such models could be used in lieu of authentic standards for the partial identification of chromatographic peaks. The results also suggested the need for further investigations into the problem of improving predictions of retention indices for such diverse sets of compounds on the more polar phases.

Similar work on the correlation between molecular structures and retention indices of different classes of compounds has been carried out by other authors.⁹⁻¹³ Different structural variables (molecular connectivity indices, variables accounting for the number and position of heteroatoms present, etc.) were used as the molecular descriptors to give excellent correlations in some cases. However, the multilinear equations obtained were highly individual for each particular class of compounds.

Previous QSPR on GC Response Factors. The GC response factor is physically more complex than the retention time. The RF is essentially a "correction" factor which measures the response of a given compound to the detecting device. For a detailed background of response factors, the reader is referred to the literature;¹⁴⁻²¹ a brief overview is presented here.

Dietz first tabulated the response factors (RF), defined by eq i, for various compounds for the FID (flame ionization detectors), using *n*-heptane as a standard with a defined RF of 1.00. In eq i, (area of compound) and (area of standard)

$$RF_{(\text{Dietz})} = \frac{(\text{area of compound})(\text{mass of standard})}{(\text{mass of compound})(\text{area of standard})} \quad (\text{i})$$

denote the areas of the respective GC peaks, and (mass of compound) and (mass of standard), the respective masses prepared as standard solutions.

To predict response factor in cases of nonavailability of pure specimens, the "effective carbon number" (ECN) method was proposed by Scanlon and Willis.²⁰ The ECN for many compounds can be calculated from the heteroatoms and functional groups in the molecule and used to obtain RF from eq ii. However, even for those classes of compounds for which the ECN can be calculated, both calculated and measured RF's can vary by 25%.

$$RF = \frac{(\text{MW of compound})(\text{ECN of standard})}{(\text{MW of standard})(\text{ECN of compound})} \quad (\text{ii})$$

Earlier papers had calculated thermal conductivity detector response factors with a light carrier gas for a series of

hydrocarbons, alcohols, ethers, and ketones using the molecular diameter approach.^{22,23} Anomalous response factors for halogenated compounds were predicted and explained by their molecular diameters which were found to be comparable with analogous hydrocarbons.²⁴ Further work in this area applied kinetic theory for the prediction of the thermal conductivity detector response.²⁵

The first published work on the prediction of GC response factors for a wide cross section of classes of organic structures, by our laboratory in collaboration with Musumarra's group,²⁶ was for 100 substituted benzenes and pyridines using a multivariate statistical partial least squares (PLS) treatment. In this PLS analysis, the Dietz response factor as the dependent variable was described as a function of 17 explanatory variables, which included the molecular weight together with structural features such as the number of atoms of each element, of multiple bonds, of functional groups, etc. A three-component PLS model explained 84% of the variance in the RF_{Dietz} data. To our knowledge, there are no other reports on computer-assisted means to predict GC response factors for a diverse set of compounds.

EXPERIMENTAL SECTION

The determination of the retention times and response factors of all standard compounds was done on a Hewlett-Packard 5890 Series II gas chromatograph equipped with a flame ionization detector (FID) and split injection port. A Hewlett-Packard fused silica capillary column (HP-1, cross-linked methyl silicone) of length 25 m and i.d. 0.32 mm was used in the split mode with the split ratio set at 50:1. The flow rates of the helium carrier gas, hydrogen, and air at room temperature (23 °C) was measured at 29, 39, and 380 mL/min, respectively. The injection temperature was 200 °C, and the following programmed oven temperature was used: 50 °C for 1 min, increasing by 20 °C/min up to 250 °C. Standard solutions were prepared for each compound and the internal standard (heptane with RF defined as 1.00) of ca. 100 mg, dissolved in either diethyl ether (50 mL), chloroform (50 mL), or methanol (50 mL). In the case of response factors, it was important for the accurate reproduction of results to ensure that the compound was completely soluble in the chosen solvent system. The volume of the sample injected was 0.5 µL with a concentration of 2 µg/mL. Compounds were obtained commercially from various sources (Aldrich, Fisher, Reilly Industries, Kodak, etc.); they were checked for purity and, where necessary, purified to at least 99%. Measured response factors were obtained from eq i.

QSPR TREATMENT: THE CODESSA PROGRAM

Our laboratory has been involved for the past five years in the development of QSPR. The work has involved three phases: in the first, Dr. K. Osmialowski and Dr. S. J. Cato collected and collated a wide variety of molecular descriptors from the literature, and in the second, Dr. E. V. Gordeeva helped produce an initial GOOUND and GROUNDSTAT program.²⁶ The third phase of our efforts has now culminated

- (9) Mihara, S.; Enomoto, N. *J. Chromatogr.* **1985**, *324*, 428.
- (10) Mihara, S.; Masuda, H. *J. Chromatogr.* **1987**, *402*, 309.
- (11) Bartle, K. D.; Lee, M. L.; Wise, S. A. *Chromatographia* **1981**, *14*, 69.
- (12) Doherty, P. J.; Hoes, R. M.; Robbat, A., Jr.; White, C. M. *Anal. Chem.* **1984**, *56*, 2697.
- (13) Robbat, A., Jr.; Xyrafas, G.; Marshall, D. *Anal. Chem.* **1988**, *60*, 982.
- (14) Sternberg, J. C.; Galloway, W. S.; Jones, D. T. L. The Mechanism of Response of Flame Ionization Detectors. In *Gas Chromatography*; Brenner, N., Callen, J. E., Weiss, M. D., Eds.; Academic Press: New York, 1962; pp 231-67.
- (15) Perkins, G., Jr.; Ronayheb, G. M.; Lively, L. D.; Hamilton, W. C. Response of the Gas Chromatographic Flame Ionization Detector to Different Functional Groups. In *Gas Chromatography*; Brenner, N., Callen, J. E., Weiss, M. D., Eds.; Academic Press: New York, 1962; pp 269-86.
- (16) Ackman, R. G. *J. Gas Chromatogr.* **1964**, *2*, 173.
- (17) Addison, R. F.; Ackman, R. G. *J. Gas Chromatogr.* **1964**, *6*, 135.
- (18) Ackman, R. G. *J. Gas Chromatogr.* **1968**, *6*, 497.
- (19) Dietz, W. A. *J. Gas Chromatogr.* **1967**, *5*, 68.
- (20) Scanlon, J. T.; Willis, D. E. *J. Chromatogr. Sci.* **1985**, *23*, 333.
- (21) Clementi, S.; Savelli, G.; Vergoni, M. *Chromatographia* **1972**, *5*, 413.

- (22) Barry, E. F.; Rosie, D. M. *J. Chromatogr.* **1971**, *59*, 269.
- (23) Barry, E. F.; Rosie, D. M. *J. Chromatogr.* **1971**, *63*, 203.
- (24) Barry, E. F.; Fischer, R. S.; Rosie, D. M. *Anal. Chem.* **1972**, *44*, 1559.
- (25) McMorris, D. W. *Anal. Chem.* **1974**, *46*, 42.
- (26) Katritzky, A. R.; Gordeeva, E. V. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835.

Table 1. List of the Compounds and Comparison of Observed Retention Time t_R Values (min) with Those Recalculated by a QSPR Treatment (Eq 5 in Table 4, Including All 152 Compounds), Deviations (Δ), and Percentage Deviations ($\% \Delta$) for Compounds in the Reference Set

no.	compound	M_w	t_R (min)			Δ	$\% \Delta$	no.	compound	M_w	t_R (min)			Δ	$\% \Delta$
			obsd	calcd							obsd	calcd			
1	heptane	100	1.78	1.96	0.18	10		76	1,4-naphthoquinone	160	7.68	7.78	0.10	1	
2	decane	142	4.59	4.49	-0.10	2		77	benzil	210	10.31	10.62	0.31	3	
3	dodecane	170	6.34	5.90	-0.44	7		78	benzaldehyde	106	3.92	4.10	0.18	5	
4	pentadecane	212	8.56	8.04	-0.52	6		79	<i>p</i> -tolualdehyde	120	5.10	4.88	-0.22	4	
5	eicosane	282	11.55	11.58	0.03	0		80	<i>trans</i> -cinnamaldehyde	132	6.63	6.16	-0.47	7	
6	cyclohexane	84	1.53	1.79	0.26	17		81	hexanenitrile	85	3.06	2.17	-0.89	29	
7	1-octene	112	2.53	3.01	0.48	19		82	benzonitrile	103	4.10	4.07	-0.03	1	
8	1-decene	140	4.45	4.60	0.15	3		83	benzyl cyanide	117	5.45	4.64	-0.81	15	
9	benzene	78	1.46	1.68	0.22	15		84	1-bromodecane	221	7.42	7.12	-0.30	4	
10	toluene	92	2.23	2.47	0.24	11		85	chlorocyclohexane	118.5	3.36	3.56	0.20	6	
11	<i>o</i> -xylene	106	3.43	3.58	0.15	5		86	bromobenzene	157	3.75	4.27	0.52	14	
12	ethylbenzene	106	3.12	3.60	0.48	15		87	iodobenzene	204	4.74	5.14	0.40	9	
13	propylbenzene	120	4.04	4.38	0.34	8		88	<i>p</i> -chlorotoluene	126.5	4.03	4.28	0.25	6	
14	allylbenzene	118	3.93	4.40	0.47	12		89	<i>p</i> -bromotoluene	171	5.24	5.19	-0.05	1	
14	anisole	108	3.59	4.24	0.65	18		90	1-bromonaphthalene	207	8.33	8.73	0.40	5	
15	mesitylene	120	4.18	3.99	-0.19	5		92	diphenyl ether	170	7.74	8.75	1.01	13	
16	α -methylstyrene	118	4.27	4.61	0.34	8		93	benzyl phenyl ether	184	8.87	9.21	0.34	4	
17	cyclohexylbenzene	160	7.24	6.90	-0.34	5		94	butyl phenyl ether	150	6.08	6.48	0.40	7	
18	hexamethylbenzene	162	8.21	6.61	-1.60	19		95	1,3,5-trimethoxybenzene	168	7.68	7.36	-0.32	4	
19	4-phenyl-1-butene	132	4.87	5.34	0.47	10		96	dioctyl sulfide	258	10.90	10.63	-0.27	3	
20	biphenyl	154	7.60	8.50	0.90	12		97	diphenyl sulfide	186	9.02	9.38	0.36	4	
21	cumene	120	3.75	4.44	0.69	18		98	phenyl disulfide	218	10.15	11.05	0.90	9	
22	indane	118	4.79	4.87	0.08	2		99	thioanisole	126	5.16	4.71	-0.45	9	
23	tetralin	132	5.92	5.60	-0.32	5		100	methyl benzyl sulfide	126	5.88	5.29	-0.59	10	
24	diphenylmethane	168	7.96	7.92	-0.04	1		101	1-methylpiperazine	99	2.98	3.38	0.40	13	
25	stilbene	180	9.79	9.53	-0.26	3		102	quinoline	129	6.45	6.60	0.15	2	
26	naphthalene	128	6.09	6.09	0.00	0		103	isoquinoline	129	6.65	6.58	-0.07	1	
27	1-benzyl-naphthalene	218	11.39	11.57	0.18	2		104	1,2,3,4-tetrahydroquinoline	133	7.17	6.48	-0.69	10	
28	1,1'-binaphthyl	254	14.14	14.47	0.33	2		105	1,2,3,4-tetrahydroisoquinoline	133	6.75	6.49	-0.26	4	
29	phenanthrene	178	10.27	9.75	-0.52	5		106	indole	117	6.81	6.71	-0.10	1	
30	anthracene	178	10.34	10.15	-0.19	2		107	benzimidazole	118	8.02	6.79	-1.23	15	
31	fluorene	166	10.28	8.71	-1.57	15		108	quinaldine	143	7.04	7.27	0.23	3	
32	bibenzyl	182	8.61	8.51	-0.10	1		109	phenanthridine	179	10.49	10.26	-0.23	2	
33	1-octanol	130	5.10	5.01	-0.09	2		110	acridine	179	10.35	10.60	0.25	2	
34	1-decanol	158	6.77	6.59	-0.18	3		111	1-methyl-2-pyridone	109	5.78	4.17	-1.61	28	
35	1-dodecanol	186	8.26	7.95	-0.31	4		112	thiophene	84	1.49	1.26	-0.23	15	
36	cyclopentanol	86	2.32	2.72	0.40	17		113	pyrrole	67	2.04	2.10	0.06	3	
37	cyclohexanol	100	3.27	3.91	0.64	19		114	dibenzofuran	168	8.60	9.42	0.82	9	
38	phenol	94	4.16	4.31	0.15	4		115	dibenzothiophene	184	10.11	10.03	-0.08	1	
39	<i>p</i> -cresol	108	5.02	5.07	0.05	1		116	benzothiazole	135	6.33	6.45	0.12	2	
40	<i>m</i> -cresol	108	5.03	5.05	0.02	0		117	3-picoline	93	2.97	2.98	0.01	0	
41	<i>o</i> -cresol	108	4.84	5.07	0.23	5		118	4-picoline	93	2.98	2.83	-0.15	5	
42	2-ethylphenol	122	5.58	5.79	0.21	4		119	2-picoline	93	2.56	3.01	0.45	17	
43	4-ethylphenol	122	5.81	5.83	0.02	0		120	2,3-lutidine	107	3.80	4.11	0.31	8	
44	4-isopropylphenol	136	6.32	6.73	0.41	6		121	4-lutidine	107	3.66	3.73	0.07	2	
45	2-isopropylphenol	136	6.11	6.67	0.56	9		122	2,6-lutidine	107	3.23	3.76	0.53	16	
46	2-isopropoxyphenol	152	5.95	7.21	1.26	21		123	4-ethylpyridine	107	3.97	3.76	-0.21	5	
47	benzyl alcohol	108	4.64	4.83	0.19	4		124	2-ethylpyridine	107	3.44	3.95	0.51	15	
48	1-naphthol	144	8.41	8.40	-0.01	0		125	3-ethylpyridine	107	3.92	3.72	-0.20	5	
49	5,6,7,8-tetrahydro-1-naphthol	148	8.02	7.75	-0.27	3		126	2,4,6-collidine	133	0.87	0.87	0.00	0	
50	1,3-propanediol	76	2.65	1.88	-0.77	29		127	3-cyanopyridine	104	4.19	4.10	-0.09	2	
51	1,4-butanediol	90	3.72	2.63	-1.09	29		128	4-cyanopyridine	104	3.89	4.02	0.13	3	
52	1,6-dihydroxynaphthalene	160	10.33	9.41	-0.92	9		129	3-pyridinecarboxaldehyde	107	4.17	4.12	-0.05	1	
53	1,7-dihydroxynaphthalene	160	10.30	9.45	-0.85	8		130	2-acetylpyridine	121	4.60	4.86	0.26	6	
54	octylamine	129	4.88	4.74	-0.14	3		131	3-acetylpyridine	121	5.20	4.89	-0.31	6	
55	decylamine	157	6.62	6.15	-0.47	7		132	2-amino-4,6-dimethylpyridine	122	5.85	5.69	-0.16	3	
56	dodecylamine	157	8.14	7.55	-0.59	7		133	4-(2-aminoethyl)pyridine	122	6.08	5.38	-0.70	12	
57	cyclohexylamine	99	3.05	3.89	0.84	28		134	ethyl pipercolinate	157	6.11	6.54	0.43	7	
58	cycloheptylamine	113	4.40	4.56	0.16	4		135	2-(2-hydroxyethyl)pyridine	123	5.63	5.26	-0.37	7	
59	di- <i>n</i> -butylamine	129	4.18	4.55	0.37	9		136	ethyl isonicotinate	151	6.04	6.29	0.25	4	
60	aniline	93	4.08	4.48	0.40	10		137	2,2'-bipyridine	156	7.98	8.14	0.16	2	
61	benzylamine	107	4.47	4.58	0.11	2		138	cyclohexanecarboxylic acid	128	5.62	5.74	0.12	2	
62	<i>o</i> -toluidine	107	5.00	5.24	0.24	5		139	nonanoic acid	158	6.76	6.66	-0.10	2	
63	<i>p</i> -toluidine	107	5.02	5.10	0.08	2		140	phenylacetic acid	136	6.54	6.62	0.08	1	
64	<i>N</i> -methylaniline	107	4.94	4.76	-0.18	4		141	4-phenylbutyric acid	164	7.86	7.95	0.09	1	
65	diphenylamine	169	9.16	9.82	0.66	7		142	<i>trans</i> -cinnamic acid	148	7.86	7.81	-0.05	1	
66	furfuryl	82	2.73	3.86	1.13	41		143	2,4,6-trimethylbenzoic acid	164	7.82	8.23	0.41	5	
67	2,4-dimethyl-3-pentanone	114	2.43	3.31	0.88	36		144	diethyl carbonate	118	2.29	2.77	0.48	21	
68	propiophenone	134	5.82	5.57	-0.25	4		145	benzyl acetate	150	5.73	6.07	0.34	6	
69	cyclopentanone	84	2.26	1.65	-0.61	27		146	phenyl benzoate	198	9.42	10.08	0.66	7	
70	cyclohexanone	98	3.25	3.02	-0.23	7		147	coumarin	146	7.87	7.41	-0.46	6	
71	cycloheptanone	112	4.42	3.92	-0.50	11		148	dihydrocoumarin	148	7.46	7.02	-0.44	6	
72	cyclohexyl phenyl ketone	188	9.14	8.98	-0.16	2		149	isochroman	134	4.13	0.83	0.03	4	
73	benzophenone	182	9.23	9.42	0.19	2n		150	methyl phenyl sulfoxide	140	7.07	5.51	-1.56	22	
74	acetophenone	120	4.93	4.97	0.04	1		151	nitrobenzene	123	5.09	5.00	-0.09	2	
75	1-tetralone	146	7.48	7.08	-0.40	5		152	<i>p</i> -nitrotoluene	137	6.24	5.67	-0.57	9	

Table 2. List of the Compounds and Comparison of Observed RF_{Dietz} Values with Those Recalculated by a QSPR Treatment (Eq 5 in Table 6, Including All 152 Compounds), Deviations (Δ), and Percentage Deviations ($\% \Delta$) for Compounds in the Reference Set

no.	compound	M_w	RF_{Dietz}				no.	compound	M_w	RF_{Dietz}			
			obsd	calcd	Δ	$\% \Delta$				obsd	calcd	Δ	$\% \Delta$
1	heptane	100	1.00	1.05	0.05	5	77	benzil	210	0.59	0.58	-0.01	2
2	decane	142	0.94	0.97	0.03	3	78	benzaldehyde	106	0.81	0.81	0.00	0
3	dodecane	170	0.96	0.92	-0.04	4	79	<i>p</i> -toaldehyde	120	0.81	0.83	0.02	2
4	pentadecane	212	0.90	0.85	-0.05	6	80	<i>trans</i> -cinnamaldehyde	132	0.72	0.80	0.08	11
5	eicosane	282	0.67	0.73	0.06	9	81	hexanenitrile	85	0.82	0.82	0.00	0
6	cyclohexane	84	1.04	1.06	0.02	2	82	benzonitrile	103	0.91	0.82	-0.09	10
7	1-octene	112	1.00	0.99	-0.01	1	83	benzyl cyanide	117	0.81	0.85	0.04	5
8	1-decene	140	0.94	0.94	0.00	0	84	1-bromodecane	221	0.60	0.57	-0.03	5
9	benzene	78	1.09	1.08	-0.01	0	85	chlorocyclohexane	118.5	0.71	0.70	-0.01	1
10	toluene	92	1.17	1.08	-0.09	8	86	bromobenzene	157	0.55	0.56	0.01	2
11	<i>o</i> -xylene	106	1.04	1.08	0.04	4	87	iodobenzene	204	0.43	0.45	0.02	5
12	ethylbenzene	106	0.93	1.08	0.15	16	88	<i>p</i> -chlorotoluene	126.5	0.78	0.75	-0.03	4
13	propylbenzene	120	1.05	1.06	0.01	1	89	<i>p</i> -bromotoluene	171	0.56	0.60	0.04	7
14	allylbenzene	118	1.04	1.03	-0.01	1	90	1-bromonaphthalene	207	0.47	0.51	0.04	9
15	mesitylene	120	1.15	1.07	-0.08	7	91	anisole	108	0.79	0.83	0.04	5
16	α -methylstyrene	118	1.03	1.03	0.00	0	92	diphenyl ether	170	0.86	0.73	-0.13	15
17	cyclohexylbenzene	160	0.96	0.99	0.03	3	93	benzyl phenyl ether	150	0.84	0.84	0.00	0
18	hexamethylbenzene	162	1.04	1.06	0.02	2	94	butyl phenyl ether	150	0.84	0.84	0.00	0
19	4-phenyl-1-butene	132	1.08	1.02	-0.06	6	95	1,3,5-trimethoxybenzene	168	0.46	0.39	-0.07	15
20	biphenyl	154	0.90	0.89	-0.01	1	96	dioctyl sulfide	258	0.55	0.58	0.03	5
21	cumene	120	1.12	1.06	-0.06	5	97	diphenyl sulfide	186	0.62	0.65	0.03	5
22	indan	118	1.13	1.03	-0.10	9	98	phenyl disulfide	218	0.54	0.59	0.05	9
23	tetralin	132	1.01	1.02	0.01	1	99	thioanisole	126	0.78	0.76	-0.02	3
24	diphenylmethane	168	0.85	0.90	0.05	6	100	benzyl methyl sulfide	126	0.74	0.76	0.02	3
25	stilbene	180	0.76	0.87	0.11	14	101	1-methylpiperazine	99	0.57	0.50	-0.07	12
26	naphthalene	128	1.07	0.94	-0.13	12	102	quinoline	129	0.82	0.76	-0.06	7
27	1-benzyl-naphthalene	218	0.89	0.78	-0.11	12	103	isoquinoline	129	0.82	0.76	-0.06	7
28	1,1'-binaphthyl	254	0.76	0.68	-0.08	11	104	1,2,3,4-tetrahydroquinoline	133	0.85	0.82	-0.03	4
29	phenanthrene	178	0.79	0.83	0.04	5	105	1,2,3,4-tetrahydroisoquinoline	133	0.85	0.83	-0.02	2
30	anthracene	178	0.76	0.83	0.07	9	106	indole	117	0.61	0.68	0.07	11
31	fluorene	166	0.86	0.86	0.00	0	107	benzimidazole	118	0.49	0.55	0.06	12
32	bibenzyl	182	0.90	0.89	-0.01	1	108	quinaldine	143	0.77	0.78	0.01	1
33	1-octanol	130	0.77	0.71	-0.06	8	109	phenanthridine	179	0.66	0.68	0.02	3
34	1-decanol	158	0.70	0.68	-0.02	3	110	acridine	179	0.68	0.69	0.01	1
35	1-dodecanol	186	0.59	0.65	0.06	10	111	1-methyl-2-pyridone	109	0.48	0.54	0.06	13
36	cyclopentanol	86	0.73	0.71	-0.02	3	112	thiophene	84	0.67	0.70	0.03	4
37	cyclohexanol	100	0.80	0.71	-0.09	11	113	pyrrole	67	0.74	0.73	-0.03	1
38	phenol	94	0.73	0.76	0.03	4	114	dibenzofuran	168	0.66	0.70	0.04	6
39	<i>p</i> -cresol	108	0.77	0.77	0.00	0	115	dibenzothiophene	184	0.71	0.63	-0.08	11
40	<i>m</i> -cresol	108	0.77	0.77	0.00	0	116	benzothiazole	135	0.63	0.60	-0.03	5
41	<i>o</i> -cresol	108	0.82	0.77	-0.05	6	117	3-picoline	93	0.86	0.84	-0.02	2
42	2-ethylphenol	122	0.82	0.78	-0.04	5	118	4-picoline	93	0.86	0.84	-0.02	2
43	4-ethylphenol	122	0.78	0.78	0.00	0	119	2-picoline	93	0.88	0.84	-0.04	5
44	4-isopropylphenol	136	0.72	0.78	0.06	8	120	2,3-lutidine	107	0.86	0.85	-0.01	1
45	2-isopropylphenol	136	0.81	0.78	-0.03	4	121	2,4-lutidine	107	0.86	0.85	-0.01	1
46	2-isopropoxyphenol	152	0.72	0.62	-0.10	14	122	2,6-lutidine	107	0.86	0.86	0.00	0
47	benzyl alcohol	108	0.86	0.80	-0.06	7	123	4-ethylpyridine	107	0.82	0.85	0.03	4
48	1-naphthol	144	0.58	0.67	0.09	16	124	2-ethylpyridine	107	0.80	0.85	0.05	6
49	5,6,7,8-tetrahydro-1-naphthol	148	0.82	0.75	-0.07	9	125	3-ethylpyridine	107	0.86	0.85	-0.01	1
50	1,3-propanediol	76	0.45	0.46	0.01	2	126	2,4,6-collidine	133	0.87	0.87	0.00	0
51	1,4-butanediol	90	0.47	0.53	0.06	13	127	3-cyanopyridine	104	0.77	0.67	-0.10	13
52	1,6-dihydroxynaphthalene	160	0.58	0.55	-0.03	5	128	4-cyanopyridine	104	0.67	0.69	0.02	3
53	1,7-dihydroxynaphthalene	160	0.50	0.56	0.06	12	129	3-pyridinecarboxaldehyde	107	0.70	0.64	-0.06	9
54	octylamine	129	0.79	0.79	0.00	0	130	2-acetylpyridine	121	0.69	0.67	-0.02	3
55	decylamine	157	0.76	0.76	0.00	0	131	3-acetylpyridine	121	0.68	0.69	0.01	1
56	dodecylamine	157	0.69	0.72	0.03	4	132	2-amino-4,6-dimethylpyridine	122	0.66	0.73	0.07	11
57	cyclohexylamine	99	0.78	0.80	0.02	3	133	4-(2-aminoethyl)pyridine	122	0.72	0.73	0.01	1
58	cycloheptylamine	113	0.79	0.79	0.00	0	134	ethyl pipercolinate	157	0.53	0.44	-0.09	17
59	di- <i>n</i> -butylamine	129	0.75	0.74	0.01	1	135	2-(2-hydroxyethyl)pyridine	123	0.60	0.67	0.07	12
60	aniline	93	0.82	0.86	0.04	5	136	ethyl isonicotinate	151	0.58	0.57	-0.01	2
61	benzylamine	107	0.85	0.90	0.05	6	137	2,2'-bipyridine	156	0.73	0.62	-0.11	15
62	<i>o</i> -toluidine	107	0.86	0.87	0.01	1	138	cyclohexanecarboxylic acid	128	0.60	0.51	-0.09	15
63	<i>p</i> -toluidine	107	0.86	0.87	0.01	1	139	nonanoic acid	158	0.52	0.51	-0.01	2
64	<i>N</i> -methylaniline	107	0.87	0.83	-0.04	5	140	phenylacetic acid	136	0.63	0.59	-0.04	6
65	diphenylamine	169	0.69	0.67	-0.02	3	141	4-phenylbutyric acid	148	0.50	0.56	0.06	12
66	furfurylamine	82	0.58	0.66	0.08	14	142	<i>trans</i> -cinnamic acid	148	0.50	0.56	0.06	12
67	2,4-dimethyl-3-pentanone	114	0.78	0.81	0.03	4	143	2,4,6-trimethylbenzoic acid	164	0.57	0.61	0.04	7
68	propiphenone	134	0.84	0.85	0.01	1	144	diethyl carbonate	118	0.40	0.51	0.11	28
69	cyclopentanone	84	0.72	0.78	0.06	8	145	benzyl acetate	150	0.74	0.71	-0.03	4
70	cyclohexanone	98	0.76	0.79	0.03	4	146	phenyl benzoate	198	0.69	0.60	-0.03	13
71	cycloheptanone	112	0.78	0.79	0.01	1	147	coumarin	146	0.65	0.61	-0.04	6
72	cyclohexyl phenyl ketone	188	0.75	0.79	0.04	5	148	dihydrocoumarin	148	0.60	0.67	0.07	12
73	benzophenone	182	0.60	0.71	0.11	18	149	isochroman	134	0.82	0.83	0.03	4
74	acetophenone	120	0.81	0.83	0.02	2	150	methyl phenyl sulfoxide	140	0.58	0.62	0.04	7
75	1-tetralone	146	0.75	0.80	0.05	7	151	nitrobenzene	123	0.63	0.63	0.00	0
76	1,4-naphthoquinone	160	0.68	0.61	-0.07	10	152	<i>p</i> -nitrotoluene	137	0.66	0.65	-0.01	2

Table 3. Quantum-Chemical and Conventional Molecular Descriptors Proposed for the Correlation of GC Retention Times t_R and Response Factors RF on the Basis of Preliminary Statistical and Chemical Analysis

no.	descriptor
Quantum-Chemical Descriptors	
1	LUMO energy
2	minimum atomic nucleophilic reactivity index for a H atom
3	minimum atomic one-electron reactivity index for a C atom
4	total hybridization component of the molecular dipole
5	maximum atomic orbital electronic population
6	maximum σ - π bond order
7	maximum bonding contribution of a MO
8	minimum valency of a H atom
9	maximum valency of a C atom
10	minimum total bond order (>0.1) of a C atom
11	minimum total bond order (>0.1) of a H atom
12	maximum total bond order of a H atom
13	maximum exchange energy for a C-H bond
14	total molecular one-center electron-nuclear attraction/number of atoms
15	total molecular one-center electron-electron repulsion/number of atoms
16	total molecular one-center electron-electron repulsion/number of atoms
17	total molecular two-center resonance energy/number of atoms
18	highest normal mode vibrational frequency
19	highest normal mode vibrational transition dipole
20	internal heat capacity of the molecule at 300 K/number of atoms
21	translational entropy of the molecule at 300 K
22	total enthalpy of the molecule at 300 K/number of atoms
23	total entropy of the molecule at 300 K/number of atoms
24	α polarizability
Conventional Descriptors	
25	molecular weight, M_w
26	number of C atoms, N_C
27	relative number of C atoms, N_C/N_A^a
28	relative weight of C atoms, $m_C N_C/M_w^b$
29	number of C-H bonds, N_{CH}
30	number of C-C bonds, N_{CC}
31	number of C-X ^c bonds, N_{CX}
32	relative number of C-H bonds, N_{CH}/N_b^d
33	relative number of C-C bonds, N_{CC}/N_b
34	relative number of C-X bonds, N_{CX}/N_b
35	number of "effective" C atoms, nc^e
36	relative number of "effective" C atoms, nc/N_A
37	relative weight of "effective" C atoms, $m_C nc/M_w$

^a N_A is the number of atoms in the molecule ^b m_C is the atomic mass of a carbon atom. ^c X is any atom. ^d N_b is the number of bonds in the molecule. ^e Number of carbon atoms in the molecule which are connected only to other carbon and/or hydrogen atoms.

in the CODESSA (comprehensive descriptors for statistical and structural analysis) program, full details of which will be published later; briefly, it utilizes over 400 conventional and quantum-chemical descriptors and operates in an MS Windows environment.

Numerous QSAR/QSPR studies successfully involve descriptors calculated from the quantum-chemical data.²⁷⁻²⁹ Thus, quantum-chemical descriptors add important supplementary information to the conventional descriptors in the search for good structure-property correlations.

The quantum-chemical descriptors were calculated by semiempirical quantum-chemical modeling of the molecules using AM1 methodology³⁰ (in the framework of the MOPAC

Table 4. Best Two- to Six-Parameter Correlations of Gas Chromatographic Retention Times t_R of 152 Compounds with a Combined Set of 37 Molecular Descriptors

eq	N^a	D^b	R^2^c	$R^2_{cv}^d$	s^e
(1)	2	8, 24	0.9134	0.9092	0.7386
(2)	3	5, 8, 24	0.9318	0.9271	0.6576
(3)	4	8, 13, 21, 24	0.9465	0.9419	0.5841
(4)	5	8, 23, 24, 25, 32	0.9539	0.9496	0.5445
(5)	6	5, 8, 23, 24, 25, 32	0.9590	0.9550	0.5152

^a Number of parameters in the correlation. ^b Descriptors involved in the correlation; numbering corresponds to that in Table 3. ^c Square of the correlation coefficient. ^d Square of the cross-validated correlation coefficient. ^e Standard deviation.

program³¹). The descriptors include the most positive and the most negative Mulliken charges on different atom types, frontier molecular orbital (FMO) energies, and respective Fukui FMO nucleophilic, electrophilic, and one-electron reactivity indices.³² The total dipole moment of the molecule, its components, and bond orders in the molecule were also used as descriptors from quantum-chemical calculations. More specific descriptors of this type are the valence-state energies of atoms and total Coulomb and exchange energies between atoms in the molecule. The principal moments of inertia of the molecule, its zero-point energy, rotational, vibrational, translational, internal, and total enthalpy, entropy, and heat capacity were also calculated by using the MOPAC program package.³¹ As solvation characteristics of polar molecules we have used the Kirkwood-Onsager solvation energy.³³

A major problem in deriving reliable structure-activity relationships from a large set of available descriptors is the selection of the restricted set of descriptors which is most appropriate for a property to be correlated and in which the collinear variables have been excluded.^{28,29}

In the CODESSA treatment, the multilinear regression analysis therefore commenced with the set of two descriptors with pair $R^2_{ij} < 0.05$, i.e. with all the orthogonal or nearly orthogonal pairs of descriptors. The best two-parameter correlation from this set was found simply by performing the treatment of the property with each of the pairs.

In the correlation treatment of the third rank (i.e. with three independent variables involved), each of the orthogonal descriptor pairs discussed above was combined with each of the other non-collinear ($R^2_{ij} < 0.65$) descriptors, and again the best correlation chosen. Similarly, in the treatment of the fourth rank (i.e. with four independent variables involved), and 400 individual variable sets of third rank with the highest R^2_{cv} (or R^2) values were used in combination with each additional non-collinear descriptor scale for a given set. Analogously, in the treatment of the fifth and higher ranks, the 400 best variable sets for the previous rank were considered in combination with each of the additional non-collinear descriptor scales.

In order to achieve the best overall correlation with the maximum information content, the number of independent variables was increased until no further statistically significant

(27) Buydens, L.; Massart, D. L.; Geerlings, P. *Anal. Chem.* **1983**, *55*, 738.

(28) Kikuchi, O. *Quant. Struct.-Act. Relat.* **1987**, *6*, 179.

(29) Cartier, A.; Rivail, J.-L. *Chemom. Intell. Lab. Syst.* **1987**, *1*, 335.

(30) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(31) Stewart, J. J. P. MOPAC Program Package 6.0. QCPE No. 455, 1990.

(32) Fukui, K. *Theory of Orientation and Stereoselection*; Springer-Verlag: Berlin, 1975.

(33) Böttcher, C. J. F. Bordewijk, P. *Theory of Electric Polarization*, 2nd ed.; Elsevier Co.: Amsterdam, 1978, Vol. 2.

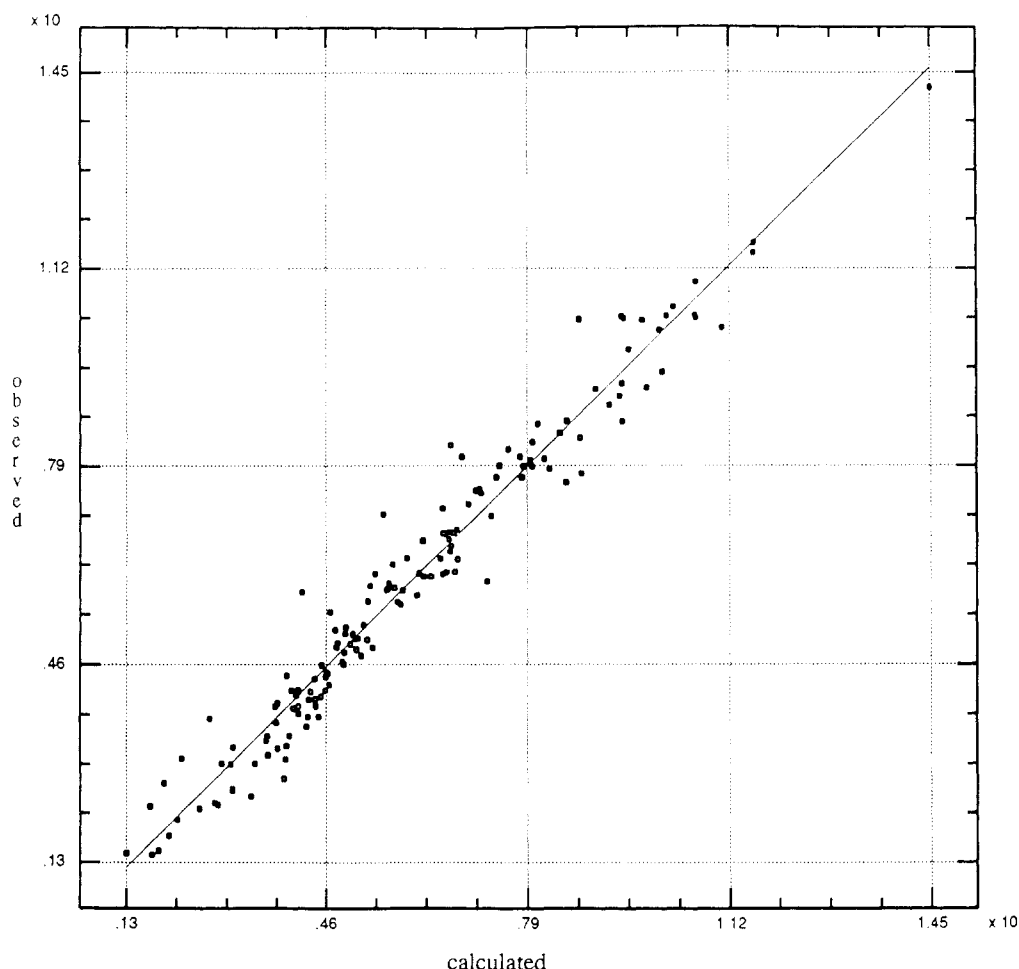


Figure 1. Correlation between the observed and predicted retention times t_R using the best six-parameter equation obtained with the combined set of 37 descriptors (eq 5 in Table 4, including all 152 compounds).

Table 5. Details of the Best Six-Parameter Correlation of the Retention Times t_R with the Combined Set of 37 Descriptors

descriptor name	coefficient ^a	error ^b	t^c	R^2_n ^d
intercept	26.50	5.98	6.942	
relative number of C-H bonds	-6.912	0.746	-9.269	0.122
total entropy of the molecule at 300 K/number of atoms	-0.871	0.102	-8.543	0.696
α polarizability	0.046	0.005	8.389	0.906
molecular weight	0.018	0.003	5.869	0.939
minimum valency of a H atom	-21.55	4.02	-5.362	0.953
maximum atomic orbital electronic population	0.929	0.218	4.256	0.959

^a Partial regression coefficient. ^b Error of the partial regression coefficient. ^c t -Value for the regression term. ^d R^2 for the correlation with this and previous terms in the table.

improvement (according to the F -ratio) of the multilinear regression was observed.

For the present statistical analysis, preliminary attempts indicated the importance of the quantum chemical descriptors. We therefore decided for both t_R and RF to select a limited number of conventional descriptors on the basis of previous experience and theory and combine those with a selection of quantum descriptors.

The initial conventional descriptor set was based on the fact that, as discussed above, the t_R and RF are strongly

Table 6. Best Two- to Six-Parameter Correlations of Gas Chromatographic Response Factors RF of 152 Compounds with a Combined Set of 37 Molecular Descriptors

eq	N^a	D^b	R^2 ^c	R^2_{cv} ^d	s^e
(1)	2	15, 28	0.7630	0.7531	0.0795
(2)	3	11, 15, 28	0.8256	0.8160	0.0684
(3)	4	10, 15, 28, 32	0.8490	0.8388	0.0639
(4)	5	8, 10, 15, 36, 37	0.8869	0.8763	0.0555
(5)	6	4, 8, 10, 15, 36, 37	0.8924	0.8810	0.0543

^a Number of parameters in the correlation. ^b Descriptors involved in the correlation; numbering corresponds to that in Table 3. ^c Square of the correlation coefficient. ^d Square of the cross-validated correlation coefficient. ^e Standard deviation.

dependent on the connectivity and other structural features of the molecule. Therefore, various *topological indices* were used by us as the input to statistical multilinear analysis. Those included the Wiener index W ,³⁴ Randic index $^1\chi$,³⁵ Kier and Hall valence connectivity index $^1\chi^v$,³⁶ Kier shape index $^1\chi$,³⁷ Kier flexibility index Φ ,³⁸ and Balaban flexibility index J .³⁹ The *geometrical descriptors* used by us reflect the three-dimensional shape of the molecule. Three shadow indices

(34) Wiener, H. J. *Am. Chem. Soc.* **1947**, *69*, 17.

(35) Randic, M. J. *Am. Chem. Soc.* **1975**, *97*, 6609.

(36) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

(37) Kier, L. B. *Acta Pharm. Jugosl.* **1986**, *36*, 171.

(38) Kier, L. B. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science Publishers: New York, 1990; pp 151-174.

(39) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.

Table 7. Details of the Best Six-Parameter Correlation of the Response Factors RF with the Combined Set of 37 Descriptors

descriptor name	coefficient ^a	error ^b	<i>t</i> ^c	<i>R</i> ² _n ^d
intercept	-2.327	0.458	-6.409	
relative weight of "effective" C atoms	-0.958	0.046	-20.762	0.465
total molecular one-center electron-electron repulsion energy	-0.0028	0.0002	-11.343	0.752
relative number of "effective" C atoms	-1.160	0.102	-11.343	0.777
minimum total bond order (>0.1) of a C atom	-0.206	0.019	-10.602	0.824
minimum valency of a H atom	3.316	0.376	8.827	0.886
total hybridization component of the molecular dipole	-0.030	0.011	-2.733	0.892

^a Partial regression coefficient. ^b Error of the partial regression coefficient. ^c *t*-Value for the regression term. ^d *R*² for the correlation with this and previous terms in the table.

Table 8. Details for the Improvement of the Correlation of RF Due to the Subsequent Addition of Important Descriptors from a Combined Set of 37 Descriptors

descriptor name	<i>R</i> ²	<i>R</i> ² _{cv}	stand deviat	<i>F</i> -ratio
(1) relative weight of "effective" C atoms	0.465 5	0.460 5	0.119 00	130.7
(2) total molecular one-center electron-electron repulsion	0.752 2	0.740 8	0.081 29	226.2
(3) relative number of "effective" C atoms	0.777 6	0.764 4	0.077 29	172.4
(4) minimum total bond order (>0.1) of a C atom	0.824 5	0.812 3	0.068 89	172.6
(5) minimum valency of an H atom	0.886 9	0.876 3	0.055 49	228.9
(6) total hybridization component of the molecular dipole	0.892 4	0.881 0	0.054 30	200.5
(7) relative number of C atoms	0.895 9	0.883 5	0.053 60	177.1
(8) minimum total bond order (>0.1) of a H atom	0.896 5	0.882 3	0.053 38	156.5
(9) maximum atomic orbital electronic population	0.898 2	0.883 0	0.053 37	139.2
(10) number of "effective" carbon atoms	0.899 7	0.882 9	0.053 16	126.5

designated *S*₁, *S*₂, and *S*₃ (areas of the molecular shadow projected on the *XY*, *YZ*, and *XZ* planes, respectively) were used in finding the correlations. Also, the normalized descriptors obtained by dividing the index by the area of the rectangle defined by the maximum lengths of the projection on the plane (*S*₄ = *S*₁/*X*_{max}*Y*_{max}, *S*₅ = *S*₂/*Y*_{max}*Z*_{max}, and *S*₆ = *S*₃/*X*_{max}*Z*_{max}) were employed.

In addition, several other structural descriptors were proposed, proceeding from two experiment-driven principles. First, the response of hydrocarbons in the FID is proportional to the carbon number of the hydrocarbons, often called the "equal-per-carbon response";¹⁴ the absolute and relative counts of carbon atoms were used. Second, the FID response of heteroatom-substituted hydrocarbons is always less than that of the parent hydrocarbon and so counts of the different C-X bonds (where X is any atom) were selected because the process

of response of organic structures in the FID begins with the thermal decomposition of such bonds.¹⁴ The molecules resulting from thermal cracking undergo "chemi-ionization" due to the strongly exothermic oxidation reactions,^{14,40} and the amount of ions formed determines the conductivity which is registered as a response.⁴¹ It has been suggested that only carbon atoms which are not already oxidized in the original molecule effectively participate in producing a response in the FID.¹⁴ Therefore, the count of the defined "effective" carbon atoms was also used as a descriptor. On the basis of the preliminary treatment of data with all the above-described descriptors, a final set of 13 descriptors was chosen for the subsequent final statistical analysis of data.

The initial set of quantum-chemical descriptors was divided into five sets according to the properties of the molecule they describe. These five sets were run separately for finding the best correlations, and the best 24 descriptors (out of a total of 144) were selected. These 24 quantum-chemical descriptors were combined with the 13 conventional descriptors, and the final multiparameter regression analysis was performed on the combined set of 37 descriptors (Table 3).

RESULTS AND DISCUSSION

The objective of the current work was to find general quantitative structure-property relationships (QSPR) for the GC retention times *t*_R and response factors RF for a large group of compounds of very diverse structures. This investigation, carried out on 152 organic structures randomly selected, yet of wide diversity in functionality, demonstrates that such relationships do exist. The 152 compounds studied included 22 aliphatics, 89 carbocyclics (59 with benzene, 10 with naphthalene and six with cyclohexane rings), and 41 heterocyclics (25 pyridines, four oxygen heterocyclics, three sulfur heterocyclics, two with piperazine, and two with quinoline rings). According to the functional groups involved, the compounds represent 31 hydroxy compounds (18 alcohols and 13 phenols), 14 ketones, 14 amines (11 primary and three secondary), 7 esters, 7 halogenides (two chloro, four bromo, and one iodo compounds), six ethers, six carboxylic acids, five sulfides, five nitriles, four aldehydes, two nitrocompounds, and one sulfoxide. The number of hydrocarbons is 32, the number of compounds containing O is 62, with N atoms is 51, and with S atoms is nine (no other elements except halogens are represented). The compounds studied and the corresponding experimental data are listed for retention times in Table 1 and for response factors in Table 2. The standard deviation for the experimentally determined *t*_R was estimated to be 4% and for the RF to be 15%. The final list of the 13 conventional descriptors and the 24 quantum chemical descriptors used in the search of the correlations is given in Table 3.

Correlation of Retention Times. The 13 conventional and 24 quantum descriptors used for searching for *t*_R correlations are given in Table 3. In Table 4, the statistical characteristics of the best two- to six-parameter correlations of the GC

(40) Rice, F. O.; Herzfeld, K. F. *J. Am. Chem. Soc.* **1934**, *56*, 284.

(41) Zweig, G.; Sherma, J., Eds. *Handbook of Chromatography*; CRC Press: Boca Raton, FL, 1972; Vol. 1.

(42) Musumarra, G.; Pisano, D.; Katritzky, A. R.; Lapucha, A. R.; Luxem, F. J.; Murugan, R.; Siskin, M.; Brons, G. *Tetrahedron Comput. Methodol.* **1989**, *2*, 17.

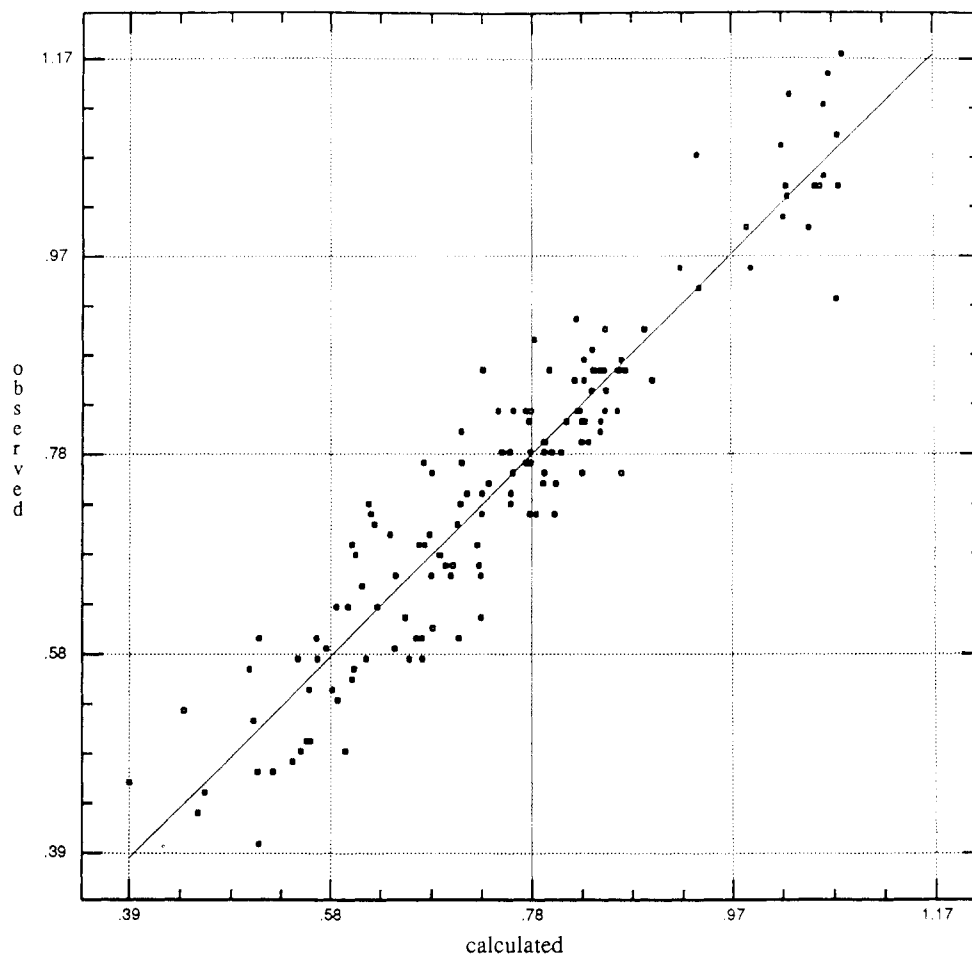


Figure 2. Correlation between the observed and predicted response factors RF using the best six-parameter equation obtained with the combined set of 37 descriptors (eq 5 in Table 6, including all 152 compounds).

retention times t_R are given. All these equations are quite successful in correlating all the retention times for the whole group of compounds, and all are highly stable as judged from their R^2 for cross-validation (R^2_{cv}) values. Figure 1 shows the correlation between the calculated and experimental retention times using the six-parameter equation. The small number of outliers shows that the six-parameter equation could be used, with considerable confidence, in the prediction of a retention time for an unknown compound. The few outliers (hexamethylbenzene, fluorene, 2-isopropoxyphenol, 1-methylpyridone and methyl phenyl sulfoxide) do not belong to any recognizable class of compounds and seem to be random. Comparisons of observed retention times t_R with those recalculated by the best obtained correlation equations, the deviations (Δ), and percentage deviations ($\% \Delta$) for compounds are presented in Table 1 and confirm this.

Physical Interpretation of Retention Time Correlation Equations. Table 5 shows the breakdown of the statistical data for the best six-parameter equation. The Student's t -values characterize the relative importance of descriptors in a particular correlation. We have also given the data on R^2 , obtained by the successive addition of descriptors into a correlation.

The most important factors in the t_R six-parameter correlation, which are also present in all correlations with lower numbers of descriptors, are the α polarizability of the molecule and the minimum valency of any H atom in it. These

quantum-chemical indices can be considered to be related to the intermolecular interaction between the molecule studied and the gas chromatographic medium. The α polarizability of the compound characterizes the effectiveness of its intermolecular induction and dispersion interaction with the medium. The positive value of the respective regression coefficient is in accordance with the physical picture—compounds with higher polarizabilities have stronger interactions with the medium and thus higher t_R values. The minimum valency of an H atom characterizes the compound as a hydrogen-bonding donor. Therefore, the presence of this term in the correlations indicates the importance of the hydrogen bond formation between the compound studied and the GC medium. The negative value of the respective regression coefficient is expected (compounds with a lower value for minimum valency have stronger hydrogen bonds and, correspondingly, longer retention times).

The other descriptors involved in the correlations are the relative number of C–H bonds and the molecular weight of the molecule. The first is approximately related to the number of CH, CH₂, and CH₃ groups in the molecule. The presence of these groups usually increases the hydrophobicity of the molecule and, thus, also the corresponding hydrophobic/hydrophilic interaction between the molecule studied and the GC medium. In accordance with the negative value of the regression coefficient for this descriptor scale, the higher hydrophobicity of the compound will lead to smaller t_R values

in a given GC medium. The correlation of t_R with the molecular weight of the compound is obvious, as larger molecules with the same homologous series have longer retention times.

Correlation of Response Factors. In the search for RF correlations, the same set of 13 conventional and 24 quantum descriptors was used (Table 3). The statistical characteristics of the best multiparameter correlations of GC response factors of 152 compounds from the preselected set of 37 descriptors are given in Table 6. The six-parameter equation ($R^2 = 0.8924$) is quite successful in correlating the GC response factors (98% of the observed values were found to be within a 95% confidence interval for predicted values), especially when considering the diversity of classes of compounds used in the study and the fact that the response factors for some classes of compounds (for example, carboxylic acids and phenols) are difficult to measure precisely.²⁶

This six-parameter equation (eq 5 in Table 6) was selected on the basis of the values of the correlation coefficient, cross-validated correlation coefficient, and the variance value (F -value). Values of these parameters for the consecutive one- to 10-parameter equations are given in Table 8. The six-parameter correlation is quite stable as judged from its R^2_{cv} (0.881) and is significant at a level less than 0.0001. Figure 2 shows the correlations between the calculated and experimental response factors using the six-parameter equation (eq 5). Thus, we believe that the six-parameter equation (eq 5) could be used with considerable confidence in the prediction of GC response factors. This is confirmed by comparisons of observed RF values with those recalculated by eq 5, which are given in Table 2.

Physical Interpretation of RF Correlations. Flame ionization is a multistep process involving the thermal decomposition of a compound with subsequent "chemi-ionization". Therefore, the yield of this process depends upon the chemical nature of the molecules and the atoms of which they are constituted. As expected from this qualitative reasoning, both types of descriptors were involved in obtaining the correlation. The most important descriptor is the relative weight of an "effective" carbon atom in the molecule, which has precedent

from the concept that only nonoxidized carbon atoms effectively produce a response in the FID. In this study, the "effective" carbon atom was defined as one connected only to either other carbon or hydrogen atoms. The relative number of "effective" carbon atoms is also important for the same reasoning. The thermal cracking of a compound inside the flame of the FID starts with the weakest C-X bond (where X is any atom).¹⁴ This being the case, the minimal total bond order of a carbon atom (defined as a minimal nondiagonal value (greater than 0.1) in the bond order matrix provided by the MOPAC program) is obviously an important descriptor. The total molecular one-center electron-electron repulsion also turned out to be an important descriptor. This quantity describes a summary of the repulsion of electrons of atoms constituting a molecule and is probably correlated with the inclination of the thermally cracked products to undergo "chemi-ionization".

CONCLUSIONS

In conclusion, we have derived significant multiparameter correlation equations in which the inclusion of quantum-chemically calculated descriptors reflects the electronic distribution and related properties of the molecules. The results obtained have been demonstrated to be useful for the prediction of GC retention times and response factors for a wide cross section of classes of organic structures. An obvious practical use is for the determination of these factors for unknown compounds present in complex mixtures. Analytical chemists thus get enhanced opportunities for performing quantitative in addition to qualitative analysis when utilizing the gas chromatograph.

ACKNOWLEDGMENT

We thank Dr. David Powell (University of Florida, Gainesville, FL) and Dr. Michael Siskin (Exxon Research & Engineering Co., Annandale, NJ) for helpful discussions.

Received for review November 19, 1993. Accepted March 8, 1994.*

* Abstract published in *Advance ACS Abstracts*, April 15, 1994.