Inter*net* Electronic Journal of Molecular Design

February 2002, Volume 1, Number 2, Pages 94–107

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Alexandru T. Balaban on the occasion of the 70th birthday
Part 2

Guest Editor: Mircea V. Diudea

Quantitative Structure-Retention Relationship Study of Gas Chromatographic Retention Indices for Halogenated Compounds

Teodora Ivanciuc[1] and Ovidiu Ivanciuc[2]

[1] University "Politehnica" of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[2] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157

Received: July 17, 2001; Accepted: October 7, 2001; Published: February 28, 2002

Citation of the article:
T. Ivanciuc and O. Ivanciuc, Quantitative Structure–Retention Relationship Study of Gas Chromatographic Retention Indices for Halogenated Compounds, *Internet Electron. J. Mol. Des.* **2002**, *1*, 94–107, http://www.biochempress.com.

Internet Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Quantitative Structure-Retention Relationship Study of Gas Chromatographic Retention Indices for Halogenated Compounds#

Teodora Ivanciuc[1] and Ovidiu Ivanciuc[2,*]

[1] University "Politehnica" of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[2] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157

Received: July 17, 2001; Accepted: October 7, 2001; Published: February 28, 2002

Internet Electron. J. Mol. Des. **2002**, *1* (2), 94–107

Abstract

Motivation. A high interest in predicting physical, chemical, and biological properties of halogenated compounds is generated by the urgent need to develop alternatives to chlorofluorocarbons, new compounds with low ozone depletion potential and a low global warming potential. Quantitative structure–retention relationships (QSRR) for halogenated compounds can be used to predict gas chromatographic retention indices by using theoretical descriptors computed from the chemical structure.

Method. QSRR models for the gas chromatographic retention indices of 207 halogenated compounds are established with the CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) program.

Results. The best results are obtained with a group of QSRR models with six structural descriptors, with a correlation coefficient between 0.994 and 0.993 and a leave–one–out cross–validation correlation coefficient between 0.992 and 0.991. All QSRR models contain the Kier and Hall connectivity index $^1\chi^v$, the number of F atoms NoF, and the gravitation index for all pairs of atoms G_1. A fourth descriptor comes from the class of charged–partial surface area indices, while the remaining two descriptors are related to the number of I or Br atoms, bond order, or Randić connectivity index $^1\chi$.

Conclusions. QSRR models developed with CODESSA allow accurate computation of the gas chromatographic retention indexes of halogenated compounds using simple constitutional, topological, geometric, electrostatic and quantum descriptors that can be computed with standard quantum chemistry packages.

Keywords. Quantitative structure–retention relationships; halogenated organic compounds; CODESSA; gas chromatographic retention indices.

1 INTRODUCTION

Halogen–containing organic compounds are an important class of chemicals, with numerous industrial and laboratory applications. They are used as solvents, plastics, anesthetics, foaming agents, refrigerants, and pesticides. Bromochlorofluorocarbons are used as fire extinguishing

Dedicated on the occasion of the 70th birthday to Professor Alexandru T. Balaban.

* Correspondence author; E–mail: ivanciuc@netscape.net.

agents. Fluoro–compounds are extensively investigated as contrast agents in X–ray and MRI, and for use as antihypoxic agents in surgical and liquid ventilation procedures. On the other hand, the use of chlorofluorocarbons (CFCs) has damaging effects on the stratospheric ozone layer, and all halogenated compounds have an important contribution in global warming by the green house effect, and halogenated organic compounds are widespread pollutants of surface and underground water. A high interest in predicting physical, chemical, and biological properties of halogenated compounds is generated by the urgent need to develop alternatives to CFCs, new compounds with low ozone depletion potential and a low global warming potential. The identification of halogenated compounds can be made with the method of gas chromatographic peak comparison with a standard sample of the each compound. Because samples of pure compounds are not always available it is important to develop quantitative structure–retention relationships (QSRR) that can efficiently predict retention parameters by using theoretical descriptors computed from the chemical structure.

Chromatographic retention is a physical phenomenon that is primarily dependent on the interactions between the solute and the stationary phase. Molecular group contribution methods are widely employed to estimate gas chromatographic retention parameters [1–3]. The difficulty of this approach is represented by the definition of a consistent set of groups and by the necessity to compute the contribution of each group from a statistically significant number of molecules where the respective group is present. This method is limited to molecules containing only the groups presented in the calibration set of molecules. Also, some group contribution schemes are not comprehensive enough to cover multiple substitutions of functional groups.

With the aid of QSRR the interactions associated with this phenomenon can be related to the constitutional, molecular graph (topological), geometrical, electrostatic, and quantum descriptors of the molecules. Gas chromatographic QSRR models have been successfully developed for a large number of compound classes: alkanes [4–6], alkenes [4,7], alkylbenzenes [8], polycyclic aromatic hydrocarbons [9], various hydrocarbons from naphthas [10], various aromatic compounds [11], alkanes, alkenes, alcohols, esters, ketones [12], monoterpenes, di– and tricyclic methyl esters and alcohols, and monocyclic ketones and alcohols [13], chlorinated alkanes [14], chlorinated benzenes [15], chlorinated dibenzodioxins [16], polychlorinated biphenyls [17,18], polyhalogenated biphenyls [19], polychlorinated dibenzofurans [18,20], pyrazines [21], diverse drug compounds [22], odor–active aliphatic compounds with oxygen–containing functional groups [23], stimulants and narcotics [24], anabolic steroids [25], sulfur vesicants [26], diverse organic compounds [27].

Integrated strategies for generating structure–property models were proposed along the years. These approaches use various programs that allow the rapid input of the chemical structures of the molecules under study, geometry optimization with molecular mechanics and quantum methods, calculation of a large set of descriptors, generation and validation of statistical structure–property models. The ADAPT program [28] is designed to compute the 3–dimensional structure of the

investigated chemical compounds, and to use constitutional, geometric and quantum information in order to generate a large number of descriptors. Various statistical models such as Multi–Linear Regression (MLR) or neural networks are tested in order to obtain the best structure–property model. ADAPT was extensively applied to the estimation of gas chromatographic retention indices of various classes of organic compounds [7,10,17,19,21–26]. Other software programs that can compute a large set of descriptors and then use them in various QSAR models are OASIS (Optimized Approach based on Structural Indices Selection) [29], PRECLAV (PRoperty Evaluation by CLAss Variables) [30], and CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) [31–33].

The CODESSA program developed by Katritzky and coworkers [31–33] implements a structure–property approach which uses about 500 structural parameters representing constitutional, topological, geometrical, electrostatic and quantum descriptors of the chemical compounds. The computed structural descriptors are used to develop models of the investigated property. CODESSA was successfully employed in structure–property studies concerning the prediction of gas chromatographic retention indices [27], normal boiling temperatures [34,35], melting temperatures for substituted benzenes [36], physical properties of substituted pyridines [37], solubility of gases and vapors in water [38], solvent polarity scale [39], polymer glass transition temperatures [40], critical micelle concentration [41], normal boiling temperatures of organic compounds containing halogen, oxygen, and sulfur [42], liquid viscosity of organic compounds [43]. The goal of the present work is to employ computer–assisted methods to develop QSRR equations relating structural features of various halogenated compounds to their experimental gas chromatographic retention indices.

2 MATERIALS AND METHODS

The QSRR models for the estimation of the gas chromatographic retention indices of various halogenated compounds are established in the following five steps: molecular structure input and generation of the files containing the chemical structures stored in a computer–readable format; quantum mechanics geometry optimization with a semi–empirical method; structural descriptors computation; structural descriptors selection; structure–retention models generation with the multi–linear regression method.

GC Retention Data. Retention indices (*RI*) of 207 halogenated aliphatic and alicyclic compounds were taken from the literature [44], and are presented in Table 1. The chromatograms were recorded on a Hewlett–Packard Model 5890A gas chromatograph equipped with flame ionization detector, with the following operating conditions: fused–silica 50 m × 0.2 mm capillary column coated with methyl silicone, carrier gas (nitrogen) flow velocity 10 cm/s, injection port and detector temperatures 270 °C. The temperature program comprised an initial isothermal period of 5

min at 40 °C, followed by a temperature increase of 2 °C/min up to 200 °C, and a final isothermal stage at this temperature. The retention indices of the halogenated compounds were computed using 1–bromoalkanes as standards. A good linear relationship between retention time and carbon number of the 1–bromoalkanes was observed, except for bromomethane and bromoethane. If the retention time of any compound was smaller than that of bromomethane, the retention index for that compound was calculated by considering methane as standard compound.

Table 1. Structure and Gas Chromatographic Retention Indices of 207 Halogenated Compounds

Nr.	Compound	*RI*
1	Bromotrichloromethane	434
2	Tetrabromomethane	726
3	Dichlorodifluoromethane	0
4	Trichlorofluoromethane	158
5	Chloropicrin	443
6	Tetrachloromethane	338
7	Tetraiodomethane	1497
8	Bromodichloromethane	374
9	Chlorodibromomethane	458
10	Bromoform	546
11	Chloroform	292
12	Iodoform	905
13	Bromochloromethane	284
14	Dibromomethane	363
15	Dichloromethane	206
16	Diiodomethane	582
17	Bromomethane	100
18	Iodomethane	200
19	1,1,1-Trichloro-2,2,3-trifluoroethane	210
20	1,1,2-Trichloro-1,2,2-trifluoroethane	211
21	Trichloroacetonitrile	351
22	Tetrachloroetylene	481
23	1,2-Difluoro-1,1,2,2-tetrachloroethane	387
24	Hexachloroethane	744
25	Tribromoethylene	637
26	Pentabromoethane	1172
27	Trichloroethylene	372
28	Trichloroacetaldehyde	382
29	Pentachloroethane	641
30	(*Z*)-1,2-Dibromoethylene	410
31	(*E*)-1,2-Dibromoethylene	438
32	1,2-Dibromo-1,1-dichloroethane	689
33	*rac*-1,2-Dibromo-1,2-dichloroethane	744
34	*meso*-1,2-Dibromo-1,2-dichloroethane	747
35	1,1,2,2-Tetrabromoethane	929
36	Chloroacetonitrile	345
37	1,1-Dichloroethylene	202
38	(*E*)-1,2-Dichloroethylene	234
39	1,1,1,2-Tetrachloroethane	513
40	1,1,2,2-Tetrachloroethane	565
41	Bromoethylene	131
42	1,1,2-Tribromoethane	687
43	Chloroethylene	50
44	Methylchloroformate	252
45	1,1,1-Trichloroethane	318
46	1,1,2-Trichloroethane	433
47	1-Bromo-2-chloroethane	390
48	1,1-Dibromoethane	406
49	1,2-Dibromoethane	465
50	1,2-Dichloroethane	315
51	Dichloromethyl methyl ether	340
52	1,1-Difluoroethane	0
53	1,2-Diiodoethane	682
54	Bromoethane	200
55	Chloroethane	110
56	Chloromethyl methyl ether	226
57	Iodoethane	284
58	Octachloropropane	1225
59	1,1,1,2,3,3,3-Heptachloropropane	1034
60	(*Z*)-1,2,3-Trichloropropene	532
61	(*E*)-1,2,3-Trichloropropene	573
62	Methyltrichloroacetate	585
63	1,1,2,3,3-Pentachloropropane	787
64	1,1,3-Trichloroacetone	613
65	2,3-Dibromopropene	532
66	1,1-Dichloropropene	327
67	(*Z*)-1,3-Dichloropropene	406
68	(*E*)-1,3-Dichloropropene	425
69	2,3-Dichloropropene	367
70	Methyl dichloroacetate	513
71	1,1,1,2-Tetrachloropropane	596
72	1,1,2,3-Tetrachloropropane	682
73	1,2,2,3-Tetrachloropropane	639
74	3-Bromo-1-propene	284
75	(*Z*)-1-Bromo-1-propene	265
76	(*E*)-1-Bromo-1-propene	275
77	Bromoacetone	433
78	Methyl bromoacetate	490
79	3-Chloro-1,2-dibromopropane	734
80	1,2,3-Tribromopropane	815
81	3-Chloro-1-propene	208
82	(*Z*)-1-Chloro-1-propene	194
83	(*E*)-1-Chloro-1-propene	203
84	Chloroacetone	310
85	Methyl chloroacetate	431
86	1,1,2-Trichloropropane	507
87	1,2,3-Trichloropropane	572
88	1-Bromo-3-chloropropane	521
89	1,2-Dibromopropane	515
90	1,3-Dibromopropane	597
91	2,2-Dibromopropane	436
92	1,2-Dichloropropane	362

Table 1. (Continued)

Nr.	Compound	*RI*	Nr.	Compound	*RI*
93	1,3-Dichloropropane	441	**149**	2,2-Dimethyl-1,3-dichloropentane	554
94	2,2-Dichloropropane	287	**150**	1-Bromopentane	500
95	1-Bromopropane	300	**151**	2-Bromopentane	451
96	2-Bromopropane	248	**152**	2-Bromo-2-methylbutane	416
97	1-Chloropropane	218	**153**	1-Chloropentane	423
98	2-Chloropropane	177	**154**	1-Chloro-3-methylbutane	389
99	Chloromethyl ethyl ether	315	**155**	2-Chloro-2-methylbutane	338
100	1-Iodopropane	385	**156**	1-Iodopentane	590
101	2-Iodopropane	337	**157**	1-Iodo-3-methylbutane	547
102	Hexachloro-1,3-butadiene	885	**158**	2-Iodo-2-methylbutane	509
103	Ethyl trifluoroacetate	209	**159**	Tetradecafluoro-2-methylpentane	0
104	*meso*-1,2,3,4-Tetrabromobutane	1195	**160**	Tetradecafluorohexane	0
105	(*Z*)-1,3-Dichloro-2-butene	493	**161**	α-BHC	1356
106	(*E*)-1,3-Dichloro-2-butene	509	**162**	β-BHC	1388
107	(*Z*)-1,4-Dichloro-2-butene	553	**163**	γ-BHC	1434
108	(*E*)-1,4-Dichloro-2-butene	573	**164**	δ-BHC	1454
109	3,4-Dichloro-1-butene	451	**165**	3-Bromocyclohexene	663
110	3-Bromo-1-butene	338	**166**	*cis*-1,4-Dichlorocyclohexane	775
111	4-Bromo-1-butene	378	**167**	*trans*-1,4-Dichlorocyclohexane	741
112	(*E*)-1-Bromo-2-butene	416	**168**	Bromocyclohexane	645
113	(*Z*)-2-Chloro-2-butene	278	**169**	Chlorocyclohexane	559
114	(*E*)-2-Chloro-2-butene	305	**170**	Iodocyclohexane	738
115	3-Chloro-1-butene	259	**171**	2,3-Dibromo-2,3-dimethylbutane	726
116	1-Chloro-2-methyl-1-propene	297	**172**	1,6-Dibromohexane	965
117	3-Chloro-2-methyl-1-propene	303	**173**	1-Bromohexane	600
118	1,2-Dibromobutane	622	**174**	1-Bromo-4-methylpentane	564
119	1,3-Dibromobutane	648	**175**	1-Chlorohexane	524
120	1,4-Dibromobutane	725	**176**	2-Chlorohexane	476
121	*rac*-2,3-Dibromobutane	581	**177**	3-Chlorohexane	476
122	*meso*-2,3-Dibromobutane	593	**178**	1-Iodohexane	692
123	1,2-Dichlorobutane	466	**179**	Tetradecafluoromethylcyclohexane	68
124	1,3-Dichlorobutane	490	**180**	Hexadecafluoroheptane	30
125	1,4-Dichlorobutane	558	**181**	Bromocycloheptane	779
126	2,2-Dichlorobutane	404	**182**	Chlorocycloheptane	699
127	2,3-Dichlorobutane	428	**183**	1-Bromoheptane	700
128	1,2-Dichloroethyl ethyl ether	526	**184**	2-Bromo-2-methylhexane	604
129	2,2'-Dichlorodiethyl ether	632	**185**	2-Bromo-2,4-dimethylpentane	565
130	1-Bromobutane	400	**186**	1-Chloroheptane	626
131	2-Bromobutane	356	**187**	2-Chloroheptane	572
132	2-Bromo-2-methylpropane	301	**188**	1-Iodoheptane	791
133	2-Bromoethyl ethyl ether	479	**189**	1,8-Dibromooctane	1170
134	1-Chlorobutane	320	**190**	1,8-Dichlorooctane	987
135	2-Chlorobutane	284	**191**	1-Bromooctane	800
136	1-Chloro-2-methylpropane	291	**192**	1-Bromo-2-ethylhexane	753
137	2-Chloro-2-methylpropane	220	**193**	1-Chlorooctane	720
138	1-Iodobutane	483	**194**	1-Iodooctane	892
139	2-Iodo-2-methylpropane	381	**195**	1-Bromononane	900
140	Hexachlorocyclopentadiene	999	**196**	1-Chlorononane	820
141	1,1,1,5,5,5-Hexafluoropentan-2,4-dione	208	**197**	1-Chloroadamantane	945
142	Bromocyclopentane	525	**198**	2-Chloroadamantane	989
143	1-Bromo-3-methyl-2-butene	517	**199**	1,10-Dibromodecane	1374
144	Chlorocyclopentane	449	**200**	1,10-Dichlorodecane	1192
145	1,4-Dibromopentane	772	**201**	1-Bromodecane	1000
146	1,5-Dibromopentane	836	**202**	1-Chlorodecane	919
147	2,4-Dibromopentane	677	**203**	1-Bromoundecane	1100
148	1,5-Dichloropentane	676	**204**	1-Chloroundecane	1019

Table 1. (Continued)

Nr.	Compound	*RI*
205	1,12-Dibromododecane	1579
206	1-Bromododecane	1200
207	1-Bromotridecane	1300

Molecular Structure Input. Three–dimensional models of the 207 halogenated compounds were generated with HyperChem [45] and optimized with the MM+ molecular mechanics program implemented in HyperChem. The HyperChem files were exported in a format suitable for the AMPAC [46] program.

Quantum Mechanics Computations. The geometry optimization was performed on a Pentium 200 MHz with the semiempirical quantum method AM1 [47] using the AMPAC 5.0 program [46]. The following set of key words were used in all quantum computations: AM1 PRECISE VECTORS BONDS PI KPOLAR ENPART.

Structural Descriptors Generation. The HyperChem structure files and the AMPAC output files were used by the CODESSA program [48] to compute a total of 518 structural descriptors. For certain compounds it was not possible to compute some descriptors and in the subsequent computations we have assumed a zero value in all such cases. CODESSA computes five classes of structural descriptors: constitutional (number of various types of atoms and bonds, number of rings, molecular weight, etc.); topological (Wiener index, Randić connectivity indices, Kier and Hall connectivity indices, Kier shape indices, information content indices, etc.); geometrical (gravitation indices, moments of inertia, molecular volume and surface area, etc.); electrostatic (when atomic charges are computed on the basis of atomic electronegativity: minimum and maximum partial charges, polarity parameter, charged partial surface area descriptors, etc.); quantum (minimum and maximum partial charges, Fukui reactivity indices, dipole moment, HOMO and LUMO energies, etc.). The definition of all descriptors used in the present study can be found in the CODESSA Reference and User's Manuals.

Structural Descriptors Selection. A selection of descriptors is accomplished with the aim to reduce the pool of descriptors by eliminating those that satisfy the following conditions: (*a*) the descriptor has a constant value for all molecules investigated; (*b*) in the monoparametric correlation with the gas chromatographic retention indices the descriptor has a squared correlation coefficient lower than 0.1; (*c*) in the monoparametric correlation the descriptor has a *t*–test value lower than 0.1; (*d*) in the monoparametric correlation the descriptor has a F–test value lower than 1; (*e*) the descriptor has an intercorrelation coefficient higher than 0.990 with another descriptor, and the second one has a higher squared correlation coefficient in the monoparametric correlation with the retention index. In this phase 298 descriptors were removed, and 220 descriptors were selected for the generation of the multiparametric correlations.

Structure–Retention Models Generation. The development of multilinear regression (MLR)

QSRR models with CODESSA comprises the following steps: (*a*) all biparametric regression equations are computed for pair of descriptors with an intercorrelation coefficient lower than 0.8; (*b*) the 10 best pairs of descriptors in biparametric correlations are selected for the development of multiparametric equations; (*c*) to an MLR model containing *n* descriptors a new descriptor is added to generate a model with *n* + 1 descriptors if the new descriptor is not significantly correlated with the previous *n* descriptors (intercorrelation coefficient lower than 0.8). Step (*c*) is repeated until MLR models with a certain maximum number of descriptors were obtained.

Table 2. Notation of the CODESSA Descriptors Involved in the QSRR Equations Presented in Table 3

Descriptor	Notation	Type[a]
Number of H atoms	NoH	C
Number of F atoms	NoF	C
Relative number of F atoms	RNoF	C
Number of Br atoms	NoBr	C
Relative number of Br atoms	RNoBr	C
Number of I atoms	NoI	C
Relative number of I atoms	RNoI	C
Gravitation index (all bonds)	G_2	C
Gravitation index (all pairs)	G_1	C
Topographic electronic index (all pairs) (electrostatic)	T_e	E
Total molecular surface area (electrostatic)	$TMSA_e$	E
Weighted PNSA (PNSA1*TMSA/1000) (electrostatic)	$WNSA1_e$	E
Difference in CPSAs (PPSA2–PNSA2) (electrostatic)	$DPSA2_e$	E
Difference in CPSAs (PPSA3–PNSA3) (electrostatic)	$DPSA3_e$	E
Fractional PNSA (PNSA3/TMSA) (electrostatic)	$FNSA3_e$	E
Weighted PNSA (PNSA3*TMSA/1000) (electrostatic)	$WNSA3_e$	E
XY Shadow	XY_s	G
Molecular volume	V	G
Molecular surface area	S	G
Final heat of formation	H_f	Q
Total molecular surface area (quantum)	$TMSA_q$	Q
Atomic charge weighted PNSA (quantum)	$PNSA3_q$	Q
Weighted PPSA (PPSA3*TMSA/1000) (quantum)	$WPSA3_q$	Q
Maximum π–π bond order	$BO_{\pi\pi\ max}$	Q
Total molecular two–center resonance energy / number of atoms	MTCRE	Q
α polarizability (computed from the dipole moment)	α	Q
Minimum one–electron reactivity index for a F atom	$OERI_{F\ min}$	Q
Minimum net atomic charge for a I atom	$Q_{I\ min}$	Q
Wiener index	W	T
Randić connectivity index (order 1)	$^1\chi$	T
Randić connectivity index (order 2)	$^2\chi$	T
Randić connectivity index (order 3)	$^3\chi$	T
Kier and Hall connectivity index (order 0)	$^0\chi^v$	T
Kier and Hall connectivity index (order 1)	$^1\chi^v$	T
Kier and Hall connectivity index (order 3)	$^3\chi^v$	T

[a] The descriptor type is coded with C for constitutional, E for electrostatic, G for geometric, Q for quantum, and T for topological

3 RESULTS AND DISCUSSION

Table 2 presents the notation of the 35 descriptors involved in the QSRR models developed for the gas chromatographic retention indices of halogenated compounds. The descriptor type is coded

with C for constitutional, E for electrostatic, G for geometric, Q for quantum, and T for topological.

Table 3. Statistical Indices r, r_{cv}, s, F, and Structural Descriptors Involved in the Multi–Linear Regression Equations Containing Up to Six Descriptors Developed to Compute the Gas Chromatographic Retention Indices of 207 Halogenated Compounds

Eq.	Descriptors	Type	r	r_{cv}	s	F
1	$^{1}\chi^{v}$	T	0.887	0.885	143	758
2	$^{0}\chi^{v}$	T	0.854	0.850	161	553
3	α	Q	0.772	0.765	197	302
4	$^{3}\chi^{v}$	T	0.751	0.744	205	265
5	$TMSA_q$	Q	0.722	0.715	215	223
6	V	G	0.705	0.693	220	202
7	XY_s	G	0.696	0.689	223	193
8	S	G	0.689	0.682	225	186
9	$TMSA_e$	E	0.684	0.677	226	180
10	$WPSA3_q$	Q	0.659	0.652	233	158
11	$^{0}\chi^{v}$ NoF	TC	0.956	0.955	91	1084
12	$^{0}\chi^{v}$ RNoF	TC	0.950	0.948	97	952
13	$^{1}\chi^{v}$ NoF	TC	0.946	0.945	101	873
14	$^{1}\chi^{v}$ H_f	TQ	0.943	0.941	104	816
15	$^{1}\chi^{v}$ RNoF	TC	0.942	0.939	105	801
16	$^{0}\chi^{v}$ H_f	TQ	0.940	0.938	106	777
17	$^{1}\chi^{v}$ W	TT	0.937	0.934	109	728
18	$^{1}\chi^{v}$ T_e	TE	0.933	0.929	112	680
19	$^{0}\chi^{v}$ $OERI_{F\ min}$	TQ	0.932	0.923	113	673
20	$^{0}\chi^{v}$ $^{2}\chi$	TT	0.930	0.925	114	653
21	$^{1}\chi^{v}$ NoF G_1	TCC	0.975	0.974	69	1314
22	$^{1}\chi^{v}$ NoF G_2	TCC	0.971	0.970	74	1130
23	$^{1}\chi^{v}$ NoF $WNSA1_e$	TCE	0.971	0.970	74	1122
24	$^{1}\chi^{v}$ NoF NoH	TCC	0.969	0.967	78	1024
25	$^{0}\chi^{v}$ NoF $^{3}\chi$	TCT	0.967	0.964	79	975
26	$^{1}\chi^{v}$ NoF G_1 $DPSA3_e$	TCCE	0.987	0.984	51	1865
27	$^{1}\chi^{v}$ NoF G_1 $WNSA3_e$	TCCE	0.987	0.984	51	1836
28	$^{1}\chi^{v}$ NoF G_1 $^{1}\chi$	TCCT	0.985	0.984	53	1704
29	$^{1}\chi^{v}$ NoF G_1 $DPSA2_e$	TCCE	0.985	0.983	54	1629
30	$^{1}\chi^{v}$ NoF G_1 $PNSA3_q$	TCCQ	0.983	0.982	57	1481
31	$^{1}\chi^{v}$ NoF G_1 $DPSA3_e$ NoBr	TCCEC	0.991	0.989	43	2093
32	$^{1}\chi^{v}$ NoF G_1 $DPSA3_e$ RNoBr	TCCEC	0.990	0.988	44	2020
33	$^{1}\chi^{v}$ NoF G_1 $WNSA3_e$ NoBr	TCCEC	0.990	0.988	44	1990
34	$^{1}\chi^{v}$ NoF G_1 $WNSA3_e$ RNoBr	TCCEC	0.990	0.988	45	1932
35	$^{1}\chi^{v}$ NoF G_1 $DPSA3_e$ NoI	TCCEC	0.990	0.988	45	1925
36	$^{1}\chi^{v}$ NoF G_1 $DPSA3_e$ NoI $^{1}\chi$	TCCECT	0.994	0.992	36	2534
37	$^{1}\chi^{v}$ NoF G_1 $DPSA3_e$ NoBr $BO_{\pi\pi\ max}$	TCCECQ	0.993	0.992	36	2467
38	$^{1}\chi^{v}$ NoF G_1 $WNSA3_e$ NoI $^{1}\chi$	TCCETC	0.993	0.992	37	2326
39	$^{1}\chi^{v}$ NoF G_1 $DPSA3_e$ NoI MTCRE	TCCETQ	0.993	0.991	37	2311
40	$^{1}\chi^{v}$ NoF G_1DPSA3_e RNoBr $BO_{\pi\pi\ max}$	TCCECQ	0.993	0.991	38	2284
41	$^{1}\chi^{v}$ NoF G_1 $FNSA3_e$ NoI $^{1}\chi$	TCCTEC	0.993	0.991	38	2267
42	$^{1}\chi^{v}$ NoF G_1 $WNSA3_e$ $Q_{I\ min}$ $^{1}\chi$	TCCETQ	0.993	0.991	38	2266
43	$^{1}\chi^{v}$ NoF G_1 $WNSA3_e$ NoBr $BO_{\pi\pi\ max}$	TCCECQ	0.993	0.991	38	2246
44	$^{1}\chi^{v}$ NoF G_1 $WNSA3_e$ RNoI $^{1}\chi$	TCCETC	0.993	0.991	38	2232
45	$^{1}\chi^{v}$ NoF G_1 $FNSA3_e$ RNoI $^{1}\chi$	TCCTEC	0.993	0.991	38	2225

The best multi–linear regression equations with up to six descriptors generated by CODESSA are presented in Table 3. For each equation we report the descriptors involved, their type, and the

statistical indices of the model, *i.e.* the correlation coefficient *r*, the leave–one–out cross–validation correlation coefficient r_{cv}, the standard deviation *s*, and the F–test value. An inspection of the best ten mono–parametric correlations reported in Table 3 Eqs. (1)–(10), shows that the Kier and Hall $^{1}\chi^{v}$ and $^{0}\chi^{v}$ indices, with $r = 0.887$ and $r = 0.854$, respectively, are by far the best descriptors surpassing the more complex electrostatic or quantum descriptors.

All best 10 bi–parametric correlations presented in Table 3 Eqs. (11)–(20) contain at least a topological index, either $^{0}\chi^{v}$ or $^{1}\chi^{v}$. The second descriptor is NoF in Eqs. (11) and (13), RNoF in Eqs. (12) and (15), heat of formation H_f in Eqs. (14) and (16), the minimum one–electron reactivity index for a F atom, $OERI_{F\ min}$, in Eq. (19), the topographic electronic index T_e in Eq. (18), or a second topological index, such as the Wiener index *W* in Eq. (17), or the Randić connectivity index $^{2}\chi$. The introduction of a second descriptor significantly improves the statistical indices of the QSRR model, such as in Eq. (11), when the use of $^{0}\chi^{v}$ together with NoF gives $r = 0.956$, $r_{cv} = 0.955$, $s = 91$, and $F = 1084$.

A further improvement of the QSRR models is obtained by using three descriptors, as can be seen from the statistics of Eqs. (21)–(25), with *r* between 0.975 and 0.967, and the leave–one–out cross–validation correlation coefficient r_{cv} between 0.974 and 0.964. The importance of the Kier and Hall connectivity indices in the QSRR modeling of the retention indices for halogenated compounds is once again indicated by their presence in this group of equations, namely $^{1}\chi^{v}$ in Eqs. (21)–(24) and $^{0}\chi^{v}$ in Eq. (25). The second descriptor is NoF for all five QSRR models with three structural descriptors, indicating the particular behavior of F compared with the other halogens. The third descriptor brings some variation in these equations, showing that similar QSRR models can be obtained with different combinations of structural descriptors.

The best QSRR models with four descriptors, presented in Table 3 Eqs. (26)–(30), show a further improvement in the statistical indices, with *r* between 0.987 and 0.983, and r_{cv} between 0.984 and 0.982. For all these QSRR models the first three descriptors are identical with those from Eq. (21), namely $^{1}\chi^{v}$, NoF, and the gravitation index for all pairs of atoms G_1. The fourth descriptor is either from the class of charged–partial surface area indices (CPSA) or the Randić connectivity index $^{1}\chi$. For the group of QSRR models with five descriptors, presented in Table 3 Eqs. (31)–(35), the statistical indices continue to improve, with *r* between 0.991 and 0.990, and r_{cv} between 0.989 and 0.988. Besides the three descriptors $^{1}\chi^{v}$, NoF, and G_1, which are common for all these equations, the fourth one is from the CPSA class ($DPSA3_e$ or $WNSA3_e$, both computed with atomic charges derived from electronegativity) and the fifth one is NoBr, RNoBr, or NoI. The QSRR models that contain $DPSA3_e$ are derived from Eq. (26), while those that contain $WNSA3_e$ are derived from Eq. (27).

For the ten QSRR models with six parameters presented in Table 3 Eqs. (36)–(45) the F index and leave–one–out cross–validation correlation coefficient have a maximum, and the *r* and *s*

statistical indices are improved when compared with those obtained for the MLR models with five structural descriptors, presented in Table 3 Eqs. (31)–(35). This group of QSRR models is very homogeneous from a statistical point of view, having r between 0.994 and 0.993, r_{cv} between 0.992 and 0.991, and standard deviation s between 36 and 38. This result clearly shows that starting with a large selection of structural descriptors it is possible to find several combinations of descriptors that provide models with similar good statistics. Moreover, owing to the errors in the experimental data, small statistical differences between QSRR equations are not significant, and it is almost impossible to select a "best" modeling equation in such cases. We present below the detailed QSRR Eqs. (36)–(45) containing each six theoretical descriptors:

$$RI = -247.4(\pm10.2) + 115.7(\pm5.2)^{1}\chi^{v} - 105.5(\pm2.5)\mathrm{NoF} + 0.06(\pm0.003)G_1 + 9.9(\pm0.7)\mathrm{DPSA3_e} + 83.0(\pm7.4)\mathrm{NoI} + 58.5(\pm5.5)^{1}\chi \quad (36)$$
$$n = 207 \quad r = 0.994 \quad r_{cv} = 0.992 \quad s = 36 \quad F = 2534 \quad \text{outliers: } \mathbf{36}, \mathbf{58}$$

$$RI = -243.7(\pm9.9) + 177.1(\pm2.7)^{1}\chi^{v} - 85.5(\pm1.6)\mathrm{NoF} + 0.07(\pm0.003)G_1 + 7.3(\pm0.5)\mathrm{DPSA3_e} - 31.6(\pm2.9)\mathrm{NoBr} + 62.7(\pm6.9)\,\mathrm{BO}_{\pi\pi\,\max} \quad (37)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.992 \quad s = 36 \quad F = 2467 \quad \text{outliers: } \mathbf{7}, \mathbf{21}, \mathbf{161}$$

$$RI = -176.9(\pm7.9) + 107.2(\pm5.2)^{1}\chi^{v} - 112.6(\pm2.6)\mathrm{NoF} + + 0.06(\pm0.003)G_1 - 30.7(\pm2.2)\mathrm{WNSA3_e} + 59.7(\pm5.7)^{1}\chi + 76.6(\pm7.6)\mathrm{NoI} \quad (38)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.992 \quad s = 37 \quad F = 2326 \quad \text{outliers: } \mathbf{36}, \mathbf{140}$$

$$RI = -504.2(\pm28.1) + 163.1(\pm2.5)^{1}\chi^{v} - 95.0(\pm2.0)\mathrm{NoF} + 0.06(\pm0.003)G_1 + 11.2(\pm0.7)\mathrm{DPSA3_e} + 79.6(\pm7.7)\mathrm{NoI} - 24.9(\pm2.7)\mathrm{MTCRE} \quad (39)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.991 \quad s = 37 \quad F = 2311 \quad \text{outliers: } \mathbf{7}, \mathbf{58}$$

$$RI = -230.3(\pm10.5) + 169.8(\pm2.8)^{1}\chi^{v} - 86.1(\pm1.7)\mathrm{NoF} + 0.07(\pm0.003)G_1 + 7.3(\pm0.5)\mathrm{DPSA3_e} - 225.6(\pm23.4)\mathrm{RNoBr} + 60.4(\pm7.2)\mathrm{BO}_{\pi\pi\,\max} \quad (40)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.991 \quad s = 38 \quad F = 2284 \quad \text{outlier: } \mathbf{7}$$

$$RI = -277.6(\pm12.9) + 120.8(\pm5.6)^{1}\chi^{v} - 108.3(\pm2.7)\mathrm{NoF} + 0.06(\pm0.003)G_1 + 78.6(\pm5.4)^{1}\chi - 2873.3(\pm214.4)\mathrm{FNSA3_e} + 84.4(\pm8.0)\mathrm{NoI} \quad (41)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.991 \quad s = 38 \quad F = 2267 \quad \text{outlier: } \mathbf{36}$$

$$RI = -175.2(\pm8.0) + 108.5(\pm5.3)^{1}\chi^{v} - 111.7(\pm2.6)\mathrm{NoF} + 0.06(\pm0.003)G_1 - 29.9(\pm2.2)\mathrm{WNSA3_e} + 59.1(\pm5.8)^{1}\chi + 1457.5(\pm149.7)\mathrm{Q_{I\,min}} \quad (42)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.991 \quad s = 38 \quad F = 2267 \quad \text{outliers: } \mathbf{36}, \mathbf{140}$$

$$RI = -191.4(\pm9.1) + 171.2(\pm2.9)^{1}\chi^{v} - 90.3(\pm1.9)\mathrm{NoF} + 0.07(\pm0.003)G_1 - 23.1(\pm1.7)\mathrm{WNSA3_e} - 30.8(\pm3.1)\mathrm{NoBr} + 61.1(\pm7.3)\mathrm{BO}_{\pi\pi\,\max} \quad (43)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.991 \quad s = 38 \quad F = 2246 \quad \text{outliers: } \mathbf{7}, \mathbf{21}, \mathbf{140}$$

$$RI = -173.4(\pm8.0) + 111.0(\pm5.3)^{1}\chi^{v} - 110.2(\pm2.6)\mathrm{NoF} + 0.06(\pm0.003)G_1 - 28.0(\pm2.2)\mathrm{WNSA3_e} + 59.1(\pm5.8)^{1}\chi + 416.2(\pm43.8)\mathrm{RNoI} \quad (44)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.991 \quad s = 38 \quad F = 2232 \quad \text{outlier: } \mathbf{36}$$

$$RI = -266.9(\pm12.6) + 123.9(\pm5.7)^{1}\chi^{v} - 106.1(\pm2.7)\mathrm{NoF} + 0.06(\pm0.003)G_1 + 76.5(\pm5.4)^{1}\chi - 2648.1(\pm206.6)\mathrm{FNSA3_e} + 469.8(\pm45.6)\mathrm{RNoI} \quad (45)$$
$$n = 207 \quad r = 0.993 \quad r_{cv} = 0.991 \quad s = 38 \quad F = 2225 \quad \text{outlier: } \mathbf{36}$$

In the above MLR equations, after each coefficient of a QSRR model we have provided its 95% confidence range. For each QSRR model with six structural descriptors we have identified the compounds that have absolute residuals greater than 3 standard deviations. Such compounds, identified for each equation with the corresponding number from Table 1, are considered to be outliers that cannot be modeled with the MLR model. In Table 4 we present the residual values for the entire group of outliers from Eqs. (36)–(45), namely compounds **7**, **21**, **36**, **58**, **140**, and **161**. A possible explanation for their large residuals is an error in the experimental *RI* or that the appropriate descriptors were not included in the QSRR models. No specific reason could be determined to explain why these six compounds act as outliers. Halogens are particularly difficult to characterize with structural descriptors, and the presence in the best QSRR models of descriptors related to halogen atoms (such as NoF, NoBr, RNoBr, NoI, RNoI, $O_{I\ min}$) clearly shows that generic descriptors are not suitable for this type of structure–property models.

Table 4. The Residual ($RI_{exp} - RI_{calc}$) for Compounds **7**, **21**, **36**, **58**, **140**, and **161** Computed with the QSRR Equations (36)–(45)

Equation	Residual					
	7	**21**	**36**	**58**	**140**	**161**
(36)	91	–30	144	113	–69	–90
(37)	170	–110	82	77	–73	–108
(38)	106	–53	153	68	–128	–59
(39)	126	–73	99	150	–109	–70
(40)	163	–109	81	83	–63	–101
(41)	108	–56	132	94	–73	–98
(42)	100	–52	153	70	–123	–56
(43)	171	–124	91	48	–115	–81
(44)	76	–51	153	75	–114	–50
(45)	71	–54	133	101	–64	–84

Since a large and diverse descriptor pool was computed with CODESSA and used in the heuristic MLR procedure for selecting significant structural descriptors, it is of interest to observe that only a small group of them were considered important in estimating the gas chromatographic retention indices of the 207 halogenated compounds from Table 1. As a consequence of this fact, Eqs. (36)–(45) are highly similar and have the origin in Eq. (21) which was obtained with $^{1}\chi^{v}$, NoF, and G_1. The fourth descriptor comes from the class of charged–partial surface area indices, namely $DPSA3_e$, $WNSA3_e$, $FNSA3_e$, all three computed with atomic charges derived from electronegativity. Similarly with the second descriptor NoF, which counts the F atoms in a compound, the fifth descriptor is related to the count of I or Br atoms (NoBr, RNoBr, NoI, RNoI) or the minimum atomic charge for a I atom, $O_{I\ min}$. The last descriptor from Eqs. (36)–(45) is either the Randić connectivity index $^{1}\chi$, the maximum π–π bond order $BO_{\pi\pi\ max}$, or the total molecular two–center resonance energy divided by the number of atoms MTCRE. It is interesting to remark the high relevance of simple structural descriptors in modeling the gas chromatographic retention indices, such as the halogen atom counts, Randić connectivity index $^{1}\chi$, and the corresponding Kier and Hall connectivity index $^{1}\chi^{v}$.

4 CONCLUSIONS

A successful application of the CODESSA software system was presented in this study for the prediction of the gas chromatographic retention indexes of 207 halogenated compounds using theoretical descriptors derived from the molecular structure. After a heuristic screening of relevant structural descriptors, we have selected a group of ten QSRR models representing multilinear regression equations with six descriptors. For these QSRR equations the F statistics and leave–one–out cross–validation correlation coefficient have a maximum, compared with those obtained for the MLR models with five and seven structural descriptors. The examination of the QSRR models produced shows that the structural descriptors encode information related to the molecular structure and the interactions that take place between the solute molecules and stationary phase during the gas chromatographic separation process. Different aspects of the molecular structure are encoded into the structural descriptors selected in the QSRR models. A close examination of the structural descriptors selected in the group of MLR equations with six descriptors shows that halogens are difficult to characterize with theoretical descriptors, because the best QSRR models contain each two descriptors related to halogen atoms, namely NoF and one of the following five descriptors related to Br or I atoms: NoBr, RNoBr, NoI, RNoI, and $O_{I\ min}$. This situation clearly shows that generic descriptors are not suitable for encoding the halogens effect in gas chromatographic separation process, and considerable efforts should be made in this direction in order to develop better structural descriptors that encode into a proper form the heteroatoms effect. The Randić connectivity index $^1\chi$ is a topological descriptor derived from the molecular graph that encodes the size and degree of branching of the molecules, but do not discriminate between carbon atoms and halogens. The Kier and Hall connectivity index $^1\chi^v$, derived from the Randić connectivity index $^1\chi$, encodes the size and branching degree and the chemical nature of various chemical species that appear in a hydrogen–depleted molecular graph. The charge–weighted surface area descriptors $DPSA3_e$, $WNSA3_e$, and $FNSA3_e$, all three computed with atomic charges derived from electronegativity, consistently appeared in several QSRR models. This was expected since molecular surface area has been shown to be an important descriptor in modeling the interactions between a compound and the stationary phase during a gas chromatographic separation.

Using the QSRR models proposed in this study, these simple theoretical descriptors can be used to estimate the gas chromatographic retention indices for new compounds in cases where standards are not available. Although the standard error of the estimations is large compared to the experimental uncertainty, the method is useful for investigating regularities in the retention indices and for not yet synthesized halogenated compounds. This is a general problem of QSRR models developed with theoretical descriptors for diverse sets of organic compounds. A significant improvement was recently made by using orthogonal descriptors [8,49], but further investigation is needed to improve the prediction of gas chromatographic retention indices.

5 REFERENCES

[1] C. T. Peng, S. F. Ding, R. L. Hua, and Z. C. Yang, Prediction of Retention Indexes. I. Structure–Retention Index Relationship on Polar Columns, *J. Chromatogr.* **1988**, *436*, 137–172.
[2] C. T. Peng, A Method for Tentative Identification of Unknown Gas Chromatographic Peaks by Retention Index, *J. Radioanal. Nucl. Chem.* **1992**, *160*, 449–460.
[3] C. T. Peng, Retrieval of Structure Information from Retention Index, *J. Chromatogr. A* **1994**, *678*, 189–200.
[4] F. Khorasheh, M. R. Gray, and M. L. Selucky, Correlation for Kováts Retention Index of C_9–C_{26} Monoalkyl and Polymethyl Alkanes and Alkenes, *J. Chromatogr.* **1989**, *481*, 1–16.
[5] N. Bošnjak, Z. Mihalić, and N. Trinajstić, Application of Topographic Indexes to Chromatographic Data: Calculation of the Retention Index of Alkanes, *J. Chromatogr.* **1991**, *540*, 430–440.
[6] R. Gautzsch and P. Zinn, List Operations on Chemical Graphs. 5. Implementation of Breadth–First Molecular Path Generation and Application in the Estimation of Retention Index Data and Boiling Points, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 791–800.
[7] R. H. Rohrbaugh and P. C. Jurs, Prediction of Gas Chromatographic Retention Indexes of Selected Olefins, *Anal. Chem.* **1985**, *57*, 2770–2773.
[8] O. Ivanciuc, T. Ivanciuc, D. J. Klein, W. A. Seitz, and A. T. Balaban, Quantitative Structure–Retention Relationships for Gas Chromatographic Retention Indices of Alkylbenzenes with Molecular Graph Descriptors, *SAR QSAR Environ. Res.* **2001**, *11*, 419–452.
[9] R. H. Rohrbaugh and P. C. Jurs, Prediction of Gas Chromatographic Retention Indexes of Polycyclic Aromatic Compounds and Nitrated Polycyclic Aromatic Compounds, *Anal. Chem.* **1986**, *58*, 1210–1212.
[10] T. F. Woloszyn and P. C. Jurs, Prediction of Gas Chromatographic Retention Data for Hydrocarbons from Naphthas, *Anal. Chem.* **1993**, *65*, 582–587.
[11] R. Gautzsch and P. Zinn, Use of Incremental Models to Estimate the Retention Indexes of Aromatic Compounds, *Chromatographia* **1996**, *43*, 163–176.
[12] Ch. Duvenbeck and P. Zinn, List Operations on Chemical Graphs. 3. Development of Vertex and Edge Models for Fitting Retention Index Data, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 211–219.
[13] Ch. Duvenbeck and P. Zinn, List Operations on Chemical Graphs. 4. Using Edge Models for Prediction of Retention Index Data. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 220–230.
[14] A. Sabljić, Calculation of Retention Indexes by Molecular Topology: Chlorinated Alkanes, *J. Chromatogr.* **1984**, *314*, 1–12.
[15] A. Sabljić, Calculation of Retention Indexes by Molecular Topology. Chlorinated Benzenes, *J. Chromatogr.* **1985**, *319*, 1–8.
[16] S. Sekušak and A. Sabljić, Calculation of Retention Indexes by Molecular Topology. III. Chlorinated Dibenzodioxins. *J. Chromatogr.* **1993**, *628*, 69–79.
[17] M. N. Hasan and P. C. Jurs, Computer–Assisted Prediction of Gas Chromatographic Retention Times of Polychlorinated Biphenyls, *Anal. Chem.* **1988**, *60*, 978–982.
[18] V. S. Ong and R. A. Hites, Relationship Between Gas Chromatographic Retention Indexes and Computer–Calculated Physical Properties of Four Compound Classes, *Anal. Chem.* **1991**, *63*, 2829–2834.
[19] M. N. Hasan and P. C. Jurs, Prediction of Gas and Liquid Chromatographic Retention Indexes of Polyhalogenated Biphenyls, *Anal. Chem.* **1990**, *62*, 2318–2323.
[20] A. Robbat, Jr. and C. Kalogeropoulos, Prediction of Gas Chromatographic Retention Indexes for Polychlorinated Dibenzofurans, *Anal. Chem.* **1990**, *62*, 2684–2688.
[21] D. T. Stanton and P. C. Jurs, Computer–Assisted Prediction of Gas Chromatographic Retention Indexes of Pyrazines, *Anal. Chem.* **1989**, *61*, 1328–1332.
[22] R. H. Rohrbaugh and P. C. Jurs, Prediction of Gas Chromatographic Retention Indexes for Diverse Drug Compounds, *Anal. Chem.* **1988**, *60*, 2249–2253.
[23] L. S. Anker, P. C. Jurs, and P. A. Edwards, Quantitative Structure–Retention Relationship Studies of Odor–Active Aliphatic Compounds with Oxygen–Containing Functional Groups, *Anal. Chem.* **1990**, *62*, 2676–2684.
[24] C. G. Georgakopoulos, J. C. Kiburis, and P. C. Jurs, Prediction of Gas Chromatographic Relative Retention Times of Stimulants and Narcotics, *Anal. Chem.* **1991**, *63*, 2021–2024.
[25] C. G. Georgakopoulos, O. G. Tsika, J. C. Kiburis, and P. C. Jurs, Prediction of Gas Chromatographic Relative Retention Times of Anabolic Steroids, *Anal. Chem.* **1991**, *63*, 2025–2028.
[26] T. F. Woloszyn and P. C. Jurs, Quantitative Structure–Retention Relationship Studies of Sulfur Vesicants, *Anal. Chem.* **1992**, *64*, 3059–3063.
[27] A. R. Katritzky, E. S. Ignatchenko, R. A. Barcock, V. S. Lobanov, and M. Karelson, Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Quantitative Structure–Property Relationship Treatment, *Anal. Chem.* **1994**, *66*, 1799–1807.

[28] A. J. Stuper and P. C. Jurs, ADAPT: A Computer System for Automated Data Analysis Using Pattern Recognition Techniques, *J. Chem. Inf. Comput. Sci*. **1976**, *16*, 99–105.
[29] O. Mekenyan, S. Karabunarliev, and D. Bonchev, The Microcomputer OASIS System for Predicting the Biological Activity of Chemical Compounds, *Computers Chem*. **1990**, *14*, 193–200.
[30] L. Tarko and O. Ivanciuc, QSAR Modeling of the Anticonvulsant Activity of Phenylacetanilides with PRECLAV (PRoperty Evaluation by CLAss Variables), *MATCH* (*Commun. Math. Comput. Chem.*) **2001**, *44*, 201–214.
[31] R. Murugan, M. P. Grendze, J. E. Toomey, Jr., A. R. Katritzky, M. Karelson, V. S. Lobanov, and P. Rachwal, Predicting Physical Properties from Molecular Structure, *CHEMTECH* **1994**, *24*, 17–23.
[32] A. R. Katritzky, V. S. Lobanov, and M. Karelson, QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure, *Chem. Soc. Rev*. **1995**, 279–287.
[33] M. Karelson, V. S. Lobanov, and A. R. Katritzky, Quantum–Chemical Descriptors in QSAR/QSPR Studies, *Chem. Rev*. **1996**, *96*, 1027–1043.
[34] A. R. Katritzky, L. Mu, V. S. Lobanov, and M. Karelson, Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics, *J. Phys. Chem*. **1996**, *100*, 10400–10407.
[35] A. R. Katritzky, V. S. Lobanov, and M. Karelson, Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure–Property Relationship, *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 28–41.
[36] A. R. Katritzky, U. Maran, M. Karelson, and V. S. Lobanov, Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach, *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 913–919.
[37] A. R. Katritzky, V. S. Lobanov, M. Karelson, R. Murugan, M. P. Grendze, and J. E. Toomey, Jr., Comprehensive Descriptors for Structural and Statistical Analysis. 1. Correlations Between Structure and Physical Properties of Substituted Pyridines, *Rev. Roum. Chim*. **1996**, *41*, 851–867.
[38] A. R. Katritzky, L. Mu, and M. Karelson, A QSPR Study of the Solubility of Gases and Vapors in Water, *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 1162–1168.
[39] A. R. Katritzky, L. Mu, and M. Karelson, QSPR Treatment of the Unified Nonspecific Solvent Polarity Scale, *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 756–761.
[40] A. R. Katritzky, P. Rachwal, K. W. Law, M. Karelson, and V. S. Lobanov, Prediction of Polymer Glass Transition Temperatures Using a General Quantitative Structure–Property Relationship Treatment, *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 879–884.
[41] P. D. T. Huibers, V. S. Lobanov, A. R. Katritzky, D. O. Shah, and M. Karelson, Prediction of Critical Micelle Concentration Using a Quantitative Structure–Property Relationship Approach. 1. Nonionic Surfactants, *Langmuir* **1996**, *12*, 1462–1470.
[42] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, Quantitative Structure–Property Relationship Study of Normal Boiling Points for Halogen–/ Oxygen–/ Sulfur–Containing Organic Compounds Using the CODESSA Program, *Tetrahedron* **1998**, *54*, 9192–9142.
[43] O. Ivanciuc, T. Ivanciuc, P.A. Filip, and D. Cabrol–Bass, Estimation of the Liquid Viscosity of Organic Compounds with a Quantitative Structure–Property Model, *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 515–524.
[44] A. Yasuhara, M. Morita, and K. Fuwa, Temperature–Programmed Retention Indexes of 221 Halogenated Organic Compounds with 1–Bromoalkanes as References, *J. Chromatogr*. **1985**, *328*, 35–48.
[45] HyperChem 4.5, Hypercube, Inc., 419 Phillip St., Waterloo, Ontario, Canada N2L 3X2; E–mail info@hyper.com; URL: http://www.hyper.com.
[46] AMPAC 5.0, Semichem, 7128 Summit, Shawnee, KS 66216, USA, E–mail: aholder@cctr.umkc.edu.
[47] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, AM1: A New General Purpose Quantum Mechanical Molecular Model, *J. Am. Chem. Soc*. **1985**, *107*, 3902–3909.
[48] CODESSA 2.13, Semichem, 7128 Summit, Shawnee, KS 66216, USA, E–mail: aholder@cctr.umkc.edu.
[49] O. Ivanciuc, T. Ivanciuc, D. Cabrol–Bass, and A. T. Balaban, Comparison of Weighting Schemes for Molecular Graph Descriptors: Application in Quantitative Structure–Retention Relationship Models for Alkylphenols in Gas–Liquid Chromatography, *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 732–743.