BioChem Press

Inter*net* Electronic Journal of
Molecular Design

Quantitative Structure–Property Relationships for the Normal Boiling Temperatures of Acyclic Carbonyl Compounds

Ovidiu Ivanciuc,[1] Teodora Ivanciuc,[2] and Alexandru T. Balaban[3]

[1] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157
[2] University "Politehnica" of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[3] Texas A&M University at Galveston, Department of Oceanography and Marine Sciences, 5007 Avenue U, Galveston, Texas 77551

Quantitative Structure–Property Relationships for the Normal Boiling Temperatures of Acyclic Carbonyl Compounds[#]

Ovidiu Ivanciuc,[1,]* Teodora Ivanciuc,[2] and Alexandru T. Balaban[3]

[1] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157
[2] University "Politehnica" of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[3] Texas A&M University at Galveston, Department of Oceanography and Marine Sciences, 5007 Avenue U, Galveston, Texas 77551

Abstract
Quantitative structure–property relationships (QSPR) models for the estimation of normal boiling temperatures for a set of 200 acyclic carbonyl compounds (containing mono– and dialdehydes, mono– and diketones, keto aldehydes, and esters of monocarboxylic acids) were established with the CODESSA program. The QSPR models developed with CODESSA allow accurate computation of the boiling temperatures of organic compounds using simple constitutional, topological, electrostatic and quantum indices that can be computed with standard quantum chemistry packages. For the group of 127 aldehydes and ketones, a good multiple linear regression equation was obtained using five theoretical descriptors, with the following statistical indices: $r = 0.990$, $r_{LOO} = 0.986$, $s = 5.3$ °C, and F = 1190. Equally good results were obtained for the group of 73 esters ($r = 0.993$, $r_{LOO} = 0.991$, $s = 4.2$ °C, and F = 906) and all 200 compounds ($r = 0.988$, $r_{LOO} = 0.987$, $s = 5.6$ °C, and F = 1628). Our results show that an improvement in the prediction of the boiling temperatures of organic compounds can be obtained by developing models for classes of structurally related compounds.

Keywords. Quantitative structure–property relationships; QSPR; boiling temperature; CODESSA; carbonyl compounds.

1 INTRODUCTION

During the last twenty years quantitative structure–property relationships (QSPR) and quantitative structure–activity relationships (QSAR) models have gained an extensive recognition in physical, organic, analytical, pharmaceutical and medicinal chemistry, biochemistry, chemical engineering and technology, toxicology, and environmental sciences. The main contributions to the widespread use of QSPR and QSAR models come from the development of novel structural descriptors and statistical equations relating various physical, chemical, and biological properties to

---

[#] Dedicated on the occasion of the 70th birthday to Professor Milan Randić.
* Correspondence author; E–mail: ivanciuc@netscape.net.

the chemical structure. The success of the QSPR and QSAR approach can be explained by the insight offered into the structural determination of chemical properties, and the possibility to estimate the properties of new chemical compounds without the need to synthesize and test them. The main hypothesis in the QSPR and QSAR approach is that all properties (physical, chemical, and biological) of a chemical substance are statistically related to its molecular structure. The investigation of large and diverse molecular data bases was made possible by the advent of general QSPR/QSAR programs [1,2], such as ADAPT [3–12], OASIS [13,14] PRECLAV [15], SciQSAR [16], and CODESSA [17–25], which integrate the computation of structural descriptors with the generation of structure–property models. These programs compute more than one thousand structural descriptors from five classes: constitutional, graph theoretic and topological indices, geometrical, electrostatic, and quantum–chemical descriptors. Using statistical methods, such as multiple linear regression (MLR), PCA, PLS, or neural networks, the best descriptors are selected in the final structure–property model.

The ability to predict with a high confidence level the physical, chemical, or biological properties for new chemicals significantly reduces the cost and time involved in the design of compounds with desired properties. Many QSPR and QSAR models were developed for the prediction of a wide range of properties, such as melting and boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, density, aqueous solubility, 1–octanol/water partition coefficient, solvation free energy, receptor binding affinities, pharmacological activities, and enzyme inhibition constants. The normal boiling temperature (t_b) of an organic compound is of high importance in the design of industrial processes, and numerous methods have been developed over the years for its estimation from the chemical structure [5–12,20–22,26–79]. Molecular group contribution methods are widely employed to estimate boiling temperatures [26–29]. 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.

In the past, the boiling temperature was mainly computed with group contribution methods, while nowadays the tendency is to employ theoretical descriptors traditionally used in QSPR and QSAR. Initial work in applying QSPR and topological indices to boiling temperature was done by Wiener [30] and Platt [31] who introduced the Wiener index *W* (defined as the sum of the distances between any two carbon atoms in the molecule) and the polarity number *p* (defined as the number of pairs of vertices separated by three edges) as sensitive structural descriptors for alkanes. Subsequently, the normal boiling temperature of alkanes was extensively used as a benchmark property in testing novel structural descriptors or QSPR models [32–48]. Since the pioneering work

of Wiener and Platt comprehensive efforts were made to apply various structural descriptors and QSPR models for the boiling temperature of an ever increasing group of homologous and congeneric series: aliphatic hydrocarbons [5,49], aromatic hydrocarbons [50–52], various classes of hydrocarbons [10], halogenated alkanes [22,53–57], acyclic ethers, peroxides, acetals, and their sulfur analogues [22,58–61], sulfides [62], alcohols [63], chlorosilanes [64], acyclic carbonyl compounds [65], nitriles [66], furans, tetrahydrofurans, thiophenes [6], pyrans, pyrroles [7], and diverse heterocyclic compounds [8]. Another tendency is to develop QSPR equations for very diverse data bases of organic compounds, with the intention to obtain boiling temperature models widely applicable, if not for all organic compounds, then for a large diversity of chemicals [9,11,12,20,21]. Both approaches are important, because QSPR models of homologous and congeneric series have a lower error in prediction, while general boiling temperature models can be used for a quick and rough estimation of this property for any organic compound. Recently, the normal boiling temperatures for a set of 200 acyclic carbonyl compounds (containing mono– and dialdehydes, mono– and diketones, keto aldehydes, and esters of monocarboxylic acids) were modeled with MLR equations [65]. In these QSPR equations, Balaban, Mills, and Basak investigated the relationship between various topological indices (computed from the molecular graph) and the boiling temperature, demonstrating that structural descriptors derived from molecular graphs can model with good accuracy this property for acyclic carbonyl compounds. In this paper we improve the boiling temperature QSPR models for this class of compounds by using a larger population of structural descriptors, as implemented in CODESSA.

2 MOLECULAR DATABASE AND QSPR METHOD

Data Base. All QSPR models are obtained with the database assembled by Balaban, Mills, and Basak [65]. The structure and normal boiling temperatures (in °C) for the set of 200 acyclic carbonyl compounds (containing mono– and dialdehydes, mono– and diketones, keto aldehydes, and esters of monocarboxylic acids) is reported in ref. 65, and we use the compounds arranged in the same order as in Table 1 from the above reference.

Previous QSPR Models. Balaban, Mills, and Basak developed three QSPR models, i.e. one for the group of ketones and aldehydes, the second one for esters, and the third one for all 200 carbonyl compounds. For the set of 127 aldehydes and ketones the QSPR model obtained with five topological indices is:

$$t_b = 210(\pm 21)J_y - 326(\pm 18)J + 251(\pm 7)s_0 + 61(\pm 12)IC_2 - 134(\pm 25)IC_1 - 160(\pm 14)$$
$$n = 127 \qquad r^2 = 0.9705 \qquad s = 6.49\ °C \qquad F = 796$$
(1)

With the same five graph descriptors a better QSPR model was obtained for the 73 esters:

$$t_b = 320(\pm 20)J_y - 418(\pm 22)J + 217(\pm 9)s_0 + 150(\pm 15)IC_2 - 281(\pm 26)IC_1 - 99(\pm 15)$$
$$n = 73 \qquad r^2 = 0.9866 \qquad s = 4.0\ °C \qquad F = 984$$
(2)

http://www.biochempress.com

Finally, by combining all compounds into a single set, the statistical indices slightly decreased, showing that a greater accuracy can be obtained by generating QSPR models for each class of compounds:

$$t_b = 302(\pm13)J_y - 372(\pm15)J + 223(\pm6)s_0 + 116(\pm91)IC_2 - 272(\pm13)IC_1 - 109(\pm11)$$
$$n = 200 \qquad r^2 = 0.9640 \qquad s = 6.93\ °C \qquad F = 1039 \tag{3}$$

The five topological indices used in the above three QSPR equations are the average distance–based connectivity index (Balaban index) J computed from the simple molecular graph (with all atoms considered as carbons) [80,81], the Balaban index J_y in which the relative covalent radius (relative to that of carbon) accounts for the presence of heteroatoms [82,83], the sum of square roots of vertex degrees s_0, and information content indices IC_1 and IC_2 [84].

Molecular Modeling. In the present investigation, the chemical structures were generated with HyperChem [85], the geometry optimization was performed with MOPAC [86] using the semiempirical quantum method AM1 [87] and the QSPR models were computed with CODESSA [88,89].

Structural Descriptors. The HyperChem structure files and the MOPAC output files were used by the CODESSA program to calculate 366 descriptors. CODESSA computes five classes of structural descriptors: constitutional (number of various types of atoms and bonds, number of rings, molecular weight); topological (Wiener index, Randić connectivity indices, Kier shape indices, information theory indices; however, till now some significant indices are not included, such as J, J_{het}, or triplet indices [32]); geometrical (principal moments of inertia, shadow indices, molecular volume and surface area); electrostatic (when atomic charges are computed on the basis of atomic electronegativity: minimum and maximum partial charges, polarity parameter, charged partial surface area descriptors, hydrogen bond donor and acceptor surface indices); quantum (minimum and maximum partial charges, Fukui reactivity indices, dipole moment, HOMO and LUMO energies, molecular polarizability, minimum/maximum valency of an atom, minimum/maximum electron–electron repulsion for an atom, minimum/maximum exchange energy for a chemical bond, minimum/maximum atomic orbital electronic population, minimum/maximum nucleus–nucleus repulsion for a chemical bond, minimum/maximum electron–nucleus attraction for a chemical bond).

Multiple Linear Regression Model. From the whole set of 366 descriptors generated with CODESSA we have discarded descriptors with a constant value for all molecules in the data set. Descriptors for which values were not available for every molecule were assigned a zero value for the missing position. In the next step the number of descriptors was reduced by eliminating those with F–test values less than 1, t–test values less than 0.1 or correlation coefficients with the boiling temperature less than 0.1; as a result of this descriptor selection procedure, 198 descriptors remained for the group of ketones and aldehydes, 155 descriptors remained for the set of esters,

while for the entire data base of carbonyl compounds the selection resulted in 196 descriptors left for subsequent correlations. CODESSA develops MLR models by a heuristic method which includes the following steps: (*a*) All quasi–orthogonal pairs of structural descriptors are selected from the initial set. Two descriptors are considered orthogonal if their intercorrelation coefficient r_{ij} is lower than 0.1. (*b*) CODESSA uses the pairs of orthogonal descriptors to compute the biparametric regression equations. The most significant 10 pairs of molecular descriptors are used in the third step. (*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 given maximum number of descriptors are obtained.

Model Validation. QSPR correlations can be observed not only because a causal relationship exists between a set of descriptors and a property, but also due to statistical bias resulting from errors in determining structural descriptors, experimental errors in measuring the property, or even due to chance alone. Model validation techniques are needed in order to distinguish between true and random correlations and to estimate the predictive power of the model. Although the QSPR equations developed with CODESSA are obtained by selection of descriptors from a large pool, several descriptor selection techniques are used in order to minimize the possibility of chance correlations. In a first step, from the initial pool of descriptors, CODESSA eliminates descriptors as indicated above, thus greatly reducing the dimensionality of the problem – that of finding a QSPR equation with a good predictive power. Then, as described in the previous section, a heuristic algorithm selects only quasi–orthogonal groups of descriptors that are tested for correlation with the boiling temperatures of carbonyl compounds. This selection algorithm ensures that the probability of obtaining a chance correlation is low, and maintains a reasonable searching time. Finally, the leave–one–out (LOO) cross–validation procedure is applied to each and every MLR equation in order to estimate the prediction power of boiling temperature QSPR.

3 RESULTS AND DISCUSSION

Table 1 presents the notation and a short description of the structural descriptors involved in the QSPR models reported in this investigation; more complete definitions of the descriptors can be found in the CODESSA manuals. The statistical results obtained in the best five monoparametric correlations are presented in Table 2. The molecular polarizability α (structural descriptors SD1) gives the best QSPR models for all three experiments: aldehydes and ketones, $r^2 = 0.9434$, $r_{LOO}^2 = 0.9413$, $s = 8.8$ °C, and F = 2085; esters, $r^2 = 0.9420$, $r_{LOO}^2 = 0.9392$, $s = 8.1$ °C, and F = 1153; all molecules, $r^2 = 0.9023$, $r_{LOO}^2 = 0.9003$, $s = 11.3$ °C, and F = 1828. When modeling separately aldehydes and ketones in one group, and esters in another group, the standard deviation is below 9 °C, while for the combined set of compounds the standard deviation increases to 11.3 °C. This

finding indicates that developing separate QSPR models for classes of structurally related compounds is the best way of improving the prediction of normal boiling temperatures of carbonyl compounds. Other important descriptors in the monoparametric models are the XY shadow index, connectivity index $^1\chi$, molecular surface area, total molecular surface area, and the number of carbon atoms (SD2 through SD6).

Table 1. Notation of the Structural Descriptors Involved in the QSPR Models for the Normal Boiling Temperature of Acyclic Carbonyl Compounds

Notation	Structural Descriptor
SD1	α, molecular polarizability (computed from the dipole moment)
SD2	XY shadow index
SD3	$^1\chi$, Randić connectivity index of order 1
SD4	MSA, molecular surface area
SD5	TMSA, total molecular surface area (quantum)
SD6	number of carbon atoms
SD7	β, molecular polarizability (computed from the dipole moment)
SD8	minimum exchange energy for a C–C bond
SD9	principal moment of inertia C / number of atoms
SD10	maximum electron–electron repulsion for a C atom
SD11	YZ shadow index
SD12	HOMO energy
SD13	CIC_2, complementary information content of order 2
SD14	minimum valency of a carbon atom
SD15	principal moment of inertia B / number of atoms
SD16	RNCS, relative negative charged surface area (quantum)
SD17	DPSA3, difference of charged surface areas (electrostatic)
SD18	DPSA3, difference of charged surface areas (quantum)
SD19	maximum atomic state energy for a carbon atom
SD20	principal moment of inertia B
SD21	PNSA3, atomic charge weighted negative surface area (electrostatic)
SD22	maximum atomic orbital electronic population
SD23	FNSA3= PNSA3/TMSA, fractional PNSA3 (quantum)
SD24	WNSA3=PNSA3×TMSA/1000, weighted PNSA3 (electrostatic)
SD25	maximum nucleus–nucleus repulsion for a C–O bond
SD26	PPSA3, atomic charge weighted negative surface area (electrostatic)
SD27	FPSA1=PPSA1/TMSA, fractional PPSA1 (quantum)
SD28	WNSA2=PNSA2×TMSA/1000, weighted PNSA2 (quantum)
SD29	WNSA2=PNSA2×TMSA/1000, weighted PNSA2 (electrostatic)
SD30	minimum electron–nucleus attraction for a C–O bond
SD31	FPSA1=PPSA1/TMSA, fractional PPSA1 (electrostatic)
SD32	PNSA2, total charge weighted negative surface area (quantum)
SD33	HACA2, hydrogen acceptor donor charged surface (quantum)
SD34	Kier flexibility index
SD35	principal moment of inertia C
SD36	maximum valency of an O atom
SD37	RNCS, relative negative charged surface area (electrostatic)
SD38	ZX Shadow / ZX Rectangle

By increasing the number of descriptors up to five the LOO correlation coefficient increases indicating that the prediction of the model steadily improves. A further increase in the number of descriptors is not warranted, since the improvement in prediction is small.

Table 2. Structural Descriptors and Statistical Indices (Calibration Correlation Coefficient *r*, Leave–One–Out Cross–Validation Correlation Coefficient r_{LOO}, Standard Deviation *s*, and Fisher Test F) in the Best Five Monoparametric QSPR Models for the Normal Boiling Temperature of Acyclic Carbonyl Compounds

SD	r^2	r_{LOO}^2	*s*	F
aldehydes and ketones				
SD1	0.9434	0.9413	8.8	2085
SD2	0.9257	0.9231	10.1	1558
SD3	0.9215	0.9187	10.4	1468
SD4	0.9134	0.9103	10.9	1319
SD5	0.9049	0.9016	11.5	1190
esters				
SD1	0.9420	0.9392	8.1	1153
SD3	0.9374	0.9336	8.4	1063
SD2	0.9304	0.9267	8.9	950
SD4	0.9210	0.9162	9.5	828
SD6	0.8994	0.8944	10.7	635
aldehydes, ketones and esters				
SD1	0.9023	0.9003	11.3	1828
SD6	0.8977	0.8957	11.6	1737
SD4	0.8773	0.8748	12.7	1416
SD2	0.8744	0.8719	12.8	1379
SD5	0.8486	0.8456	14.1	1110

Table 3. Structural Descriptors and Statistical Indices in the Best Ten QSPR Models with Five Descriptors for the Normal Boiling Temperature of Acyclic Carbonyl Compounds

|		SD			r^2	r_{LOO}^2	*s*	F
aldehydes and ketones								
SD3	SD7	SD8	SD9	SD10	0.9801	0.9772	5.3	1190
SD1	SD11	SD9	SD12	SD13	0.9796	0.9764	5.4	1163
SD3	SD7	SD8	SD9	SD14	0.9795	0.9748	5.4	1155
SD3	SD7	SD8	SD15	SD14	0.9791	0.9761	5.5	1134
SD3	SD7	SD8	SD16	SD17	0.9791	0.9770	5.5	1134
SD3	SD7	SD8	SD16	SD18	0.9790	0.9769	5.5	1129
SD3	SD7	SD8	SD16	SD11	0.9789	0.9765	5.5	1121
SD3	SD7	SD8	SD15	SD10	0.9788	0.9767	5.5	1120
SD3	SD7	SD8	SD9	SD19	0.9788	0.9751	5.5	1119
SD1	SD11	SD20	SD12	SD13	0.9788	0.9758	5.5	1118
esters								
SD3	SD7	SD21	SD22	SD23	0.9854	0.9823	4.2	906
SD3	SD7	SD24	SD25	SD26	0.9853	0.9804	4.2	896
SD3	SD7	SD21	SD22	SD27	0.9852	0.9819	4.2	894
SD3	SD7	SD24	SD22	SD28	0.9850	0.9813	4.2	877
SD3	SD7	SD21	SD25	SD26	0.9848	0.9803	4.3	868
SD3	SD7	SD24	SD22	SD29	0.9848	0.9812	4.3	867
SD1	SD38	SD30	SD14	SD26	0.9847	0.9806	4.3	862
SD3	SD7	SD21	SD22	SD31	0.9847	0.9812	4.3	861
SD3	SD7	SD24	SD22	SD32	0.9846	0.9808	4.3	859
SD3	SD7	SD24	SD25	SD30	0.9846	0.9806	4.3	858
aldehydes, ketones and esters								
SD1	SD14	SD34	SD33	SD35	0.9767	0.9750	5.6	1628
SD1	SD14	SD34	SD36	SD35	0.9766	0.9749	5.6	1617
SD1	SD14	SD34	SD37	SD35	0.9761	0.9734	5.6	1585
SD1	SD14	SD34	SD33	SD20	0.9758	0.9741	5.7	1567
SD1	SD14	SD34	SD36	SD20	0.9754	0.9737	5.7	1539
SD1	SD14	SD34	SD33	SD9	0.9752	0.9736	5.8	1526
SD1	SD14	SD34	SD33	SD15	0.9751	0.9736	5.8	1523
SD1	SD14	SD34	SD36	SD9	0.9748	0.9731	5.8	1500
SD1	SD14	SD34	SD36	SD15	0.9746	0.9730	5.8	1491
SD1	SD14	SD34	SD37	SD20	0.9742	0.9720	5.9	1464

For the group of 127 aldehydes and ketones the results from Table 3 clearly indicate that the following QSPR model gives the best predictions:

$$t_b = -400.1\,(\pm 45.5) + 41.060(\pm 0.781)SD3 + 6.772 \times 10^{-2}(\pm 9.110 \times 10^{-3})SD7$$
$$+ 31.27\,(\pm 4.30)SD8 - 958(\pm 132)SD9 + 2.826(\pm 0.524)SD10 \qquad (4)$$
$$n = 127 \qquad r^2 = 0.9801 \qquad r_{LOO}^2 = 0.9772 \qquad s = 5.3\,°C \qquad F = 1190$$
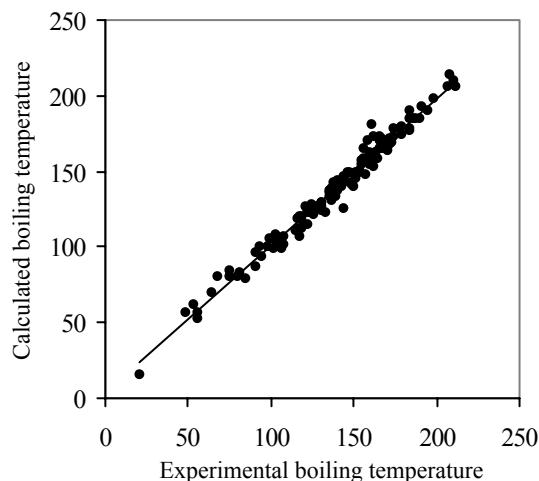
Figure 1. Experimental normal boiling temperatures vs. calculated with Eq. (4) for the set of 127 aldehydes and ketones.
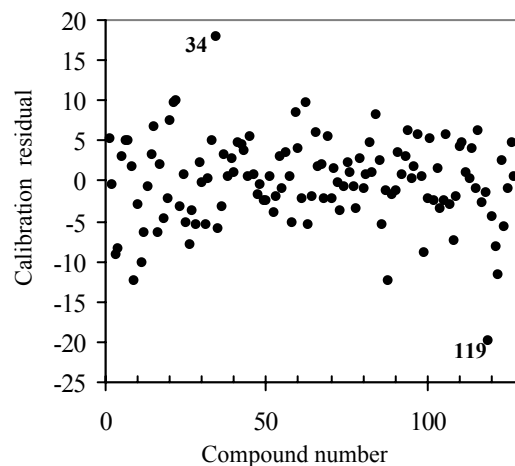
Figure 2. Calibration residuals computed with Eq. (4) for the normal boiling temperatures of 127 aldehydes and ketones.
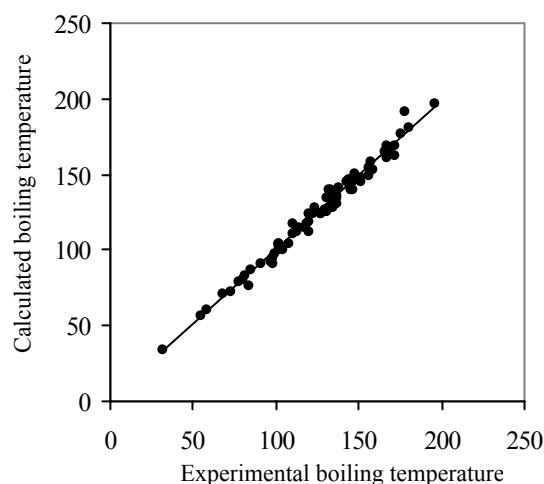
In Figure 1 we present the experimental vs. calculated boiling temperatures for the group of 127 aldehydes and ketones, while in Figure 2 we display the calibration residuals computed with Eq. (4). Both these figures show that there is no special trend of the residuals and no clusters can be detected in the data. The group of compounds with large residuals will be discussed at the end of this section. Compared with Eq. (1), obtained only with topological indices, the QSPR model from Eq. (4) has a lower standard deviation and a higher correlation coefficient, indicating that the addition of the quantum descriptors can improve the boiling temperature prediction for aldehydes and ketones. The following five theoretical descriptors are present in Eq. (4): SD3, the Randić connectivity index of order 1 $^1\chi$; SD7, the molecular polarizability β computed from the dipole moment; SD8, minimum exchange energy for a C–C bond; SD9, principal moment of inertia C / number of atoms; SD10, maximum electron–electron repulsion for a C atom. From these five descriptors, three (namely, SD3, SD7, and SD8) are present in eight out of ten QSPR models for the group of 127 aldehydes and ketones reported in Table 3, indicating that this set of descriptors is important in predicting the boiling temperatures for this group of organic compounds. An examination of the ten QSPR models for aldehydes and ketones from Table 3 reveals that the statistical indices are very close, and all equations have similar predictive power. In this context it is fit to recall that QSPR equations represent statistical models between a group of independent variables and a group of dependent variables. Although such models can be used for making

http://www.biochempress.com

predictions for new compounds, in order to give insight into the mechanism of action of chemicals, or to suggest important descriptors that determine a given property, we always have to consider that QSPR models are not causal but statistical, and therefore a descriptor can be selected not only due to its relationship with the investigated property, but also by chance alone. Moreover, structural descriptors can be intercorrelated, and in such cases similar statistics for QSPR models can be obtained with different sets of descriptors. The prediction experiments performed with the leave–one–out cross–validation procedure show that r_{LOO} is very close to the calibration correlation coefficient r, demonstrating that this QSPR equation has a good prediction power.

For the group of 73 esters the QSPR results presented in Table 3 imply that the statistical quality of all ten equations is very similar, with best results being offered by the following model:

$$t_b = -13286(\pm1710) + 40.72(\pm1.21)SD3 + 8.248\times10^{-2}(\pm8.56\times10^{-3})SD7 - 5.594(\pm0.589)SD21$$
$$+ 6917(\pm893)SD22 + 269.1(\pm70.7)SD23 \qquad (5)$$
$$n = 73 \qquad r^2 = 0.9854 \qquad r_{LOO}{}^2 = 0.9823 \qquad s = 4.2\ °C \qquad F = 906$$

Figure 3. Experimental normal boiling temperatures vs. calculated with Eq. (5) for the set of 73 esters.
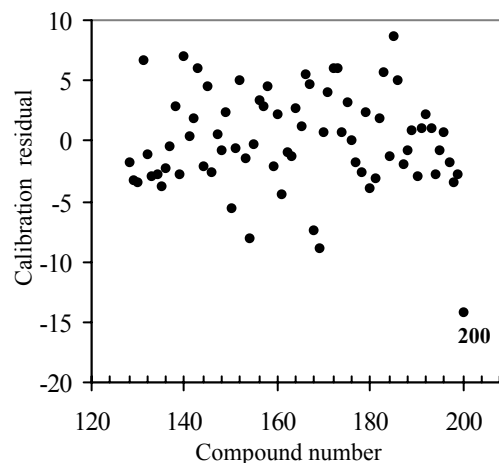
Figure 4. Calibration residuals computed with Eq. (5) for the normal boiling temperatures of 73 esters.

Using the QSPR model from Eq. (5) we present in Figure 3 the experimental vs. calculated boiling temperatures for the set of 73 esters, while in Figure 4 we display the calibration residuals. With the exception of compound 200 that has the largest residual, all other boiling temperatures are computed with a fairly good precision. Although this QSPR model was obtained by selecting structural descriptors from different classes, including geometric and quantum indices, Eq. (5) has a standard deviation larger with 0.2 °C than that of Eq. (2). Because the QSPR model from Eq. (2) contains only topological indices, it appears that our investigation did not include those geometric and quantum indices that can improve the boiling temperature modeling of esters. The following five theoretical descriptors are present in Eq. (5): SD3, the Randić connectivity index of order 1 $^1\chi$; SD7, the molecular polarizability β computed from the dipole moment; SD21, the atomic charge weighted negative surface area, PNSA3, computed with electrostatic atomic charges; SD22, the

maximum atomic orbital electronic population; SD23, the fractional atomic charge weighted negative surface area, FNSA3, computed with quantum atomic charges. From these five structural descriptors the first two are common with Eq. (4) obtained for aldehydes and ketones. SD21 and SD23 belong to the group of charged partial surface area (CPSA) [90] descriptors, defined by Jurs in terms of the solvent–accessible surface area of each atom and the atomic charge computed from the atomic electronegativity or with a quantum chemistry method. The molecule is considered as an ensemble of hard spheres defined by the van der Waals radii of the atoms. The solvent–accessible surface area is traced out by the center of a solvent sphere (usually water) that rolls over the van der Waals surface of the molecule. The CPSA descriptors encode features responsible for polar interactions between molecules. The atomic charge weighted negative surface area index PNSA3 is computed for all negatively charged atoms in the molecule:

$$PNSA3 = \sum_i SA_i^- Q_i^-$$ (6)

where SA_i^- is the surface area of the negatively charged atom i, and Q_i^- is the partial negative charge of atom i. The fractional atomic charge weighted negative surface area FNSA3 is obtained from PNSA3:

$$FNSA3 = PNSA3 / TMSA$$ (7)

where the total molecular surface area TMSA is the sum of all atomic surface areas SA_i:

$$TMSA = \sum_{i=1}^{N} SA_i$$ (8)

The CPSA indices describe in a quantitative way the interactions between polar regions of molecules; the importance of such indices for the modeling of the boiling temperature of carbonyl compounds is indicated by the presence of 14 CPSA indices among the 37 descriptors from Table 1.

The boiling temperature of the consolidated group of aldehydes, ketones, and esters is best modeled with the QSPR model:

$$t_b = -1641(\pm167) + 1.8030(\pm0.0588)SD1 + 430(\pm44.0)SD14 + 6.396(\pm0.497)SD33$$
$$+ 5.298(\pm0.421)SD34 - 135(\pm20.8)SD35$$ (9)
$$n = 200 \qquad r^2 = 0.9767 \qquad r_{LOO}^2 = 0.9750 \qquad s = 5.6\ °C \qquad F = 1628$$

In Figure 5 we present the experimental vs. calculated boiling temperatures for the 200 carbonyl compounds, while in Figure 6 we display the calibration residuals computed with Eq. (9). The above QSPR model, with $s = 5.6$ °C, represents a significant improvement compared to Eq. (3), with $s = 6.93$ °C. The following five theoretical descriptors are present in Eq. (8): SD1, the molecular polarizability α, computed from the dipole moment; SD14, the minimum valency of a carbon atom; SD33, the hydrogen acceptor donor charged surface, HACA2, computed with quantum atomic charges; SD34, the Kier flexibility index; SD35, the principal moment of inertia C. The hydrogen–bonding ability of compounds can be characterized by the index HACA2, the

hydrogen bonding acceptor charged surface area:

$$HACA2 = \sum \frac{Q_a SA_a^{1/2}}{TMSA^{1/2}} \tag{10}$$

where SA_a is the surface area of the hydrogen acceptor atom, and Q_a is the partial charge on the hydrogen bonding acceptor atom. In general, the following atoms are considered as possible hydrogen acceptors: carbonyl oxygen atoms (except in COOR), hydroxy oxygen atoms, amino nitrogen atoms, aromatic nitrogens, and mercapto sulfur atoms. The results reported in Table 3 reveal that all ten QSPR models for the 200 carbonyl compounds have similar statistical indices and predictive power; this is not surprising, since three descriptors, namely SD1, SD14, and SD34, are present in all ten equations.
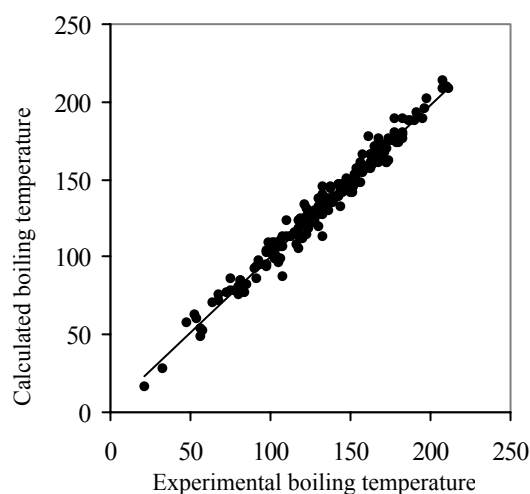
Figure 5. Experimental normal boiling temperatures vs. calculated with Eq. (9) for the set of 200 acyclic carbonyl compounds.
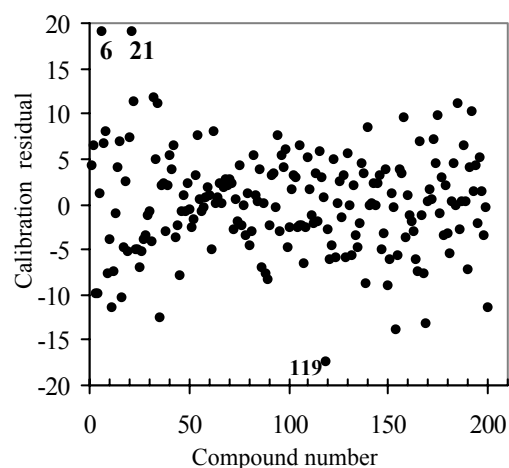
Figure 6. Calibration residuals computed with Eq. (9) for the normal boiling temperatures of 200 acyclic carbonyl compounds.

We now turn our attention to the cases of compounds with large errors in the computed boiling temperatures, since this analysis can indicate the limits of the QSPR models, or structural features that are not accurately encoded by the set of descriptors used in this study, or even possibly erroneous entries for experimental t_b. In Table 4 we have collected all carbonyl compounds that have residuals greater than 2s in one or more QSPR models; we present their label taken from Table 1 of Ref. 65, SMILES codes, experimental boiling temperatures, and residuals ($t_{b,exp} - t_{b,calc}$). The structures of these carbonyl compounds with large errors in the computed boiling temperature are presented in Figure 7. The QSPR model from Eq. (3) gives ten compounds with an absolute residual between 2s and 3s (compounds 3, 7, 9, 21, 33, 35,62, 88, 119, and 199) and two statistical outliers, with absolute residuals greater than 3s (compounds 1 and 15). In column five of Table 4 we present the residuals computed with Eq. (4) for aldehydes and ketones and with Eq. (5) for esters. For the group of 127 aldehydes and ketones we have three molecules with an absolute residual between 2s

and 3s (compounds 9, 88, and 122) and two statistical outliers (compounds 34 and 119); the two outliers are highlighted in Figure 2. Using the QSPR model from Eq. (5) developed for the 73 esters, one identifies two molecules with an absolute residual between 2s and 3s (compounds 169 and 185) and one statistical outlier (compound 200), indicated in Figure 4. An inspection of the boiling temperatures computed with Eq. (9) for all 200 compounds, helps us to find eight molecules with an absolute residual between 2s and 3s (compounds 11, 22, 32, 35, 154, 169, 185, and 200) and three statistical outliers (compounds 6, 21, and 119); these outliers are identified in Figure 6.

Table 4. Selected Carbonyl Compounds with Their Label from Table 1 of Ref. 65, SMILES Code, Experimental Boiling Temperature, and Residuals ($t_{b,exp} - t_{b,calc}$)

| No.	SMILES[a]	t_b (°C)	res. A[b]	res. B[c]	res. C[d]
1	CC=O	21	–24	5.3	4.3
3	C=CC=O	53	–20	–9.2	–9.9
6	CC#CC=O	107	3	5.0	19.1
7	CC(=O)C#C	84	17	5.2	6.7
9	CC(=C)C=O	68	–15	–12.3	–7.7
11	CCCC=O	75	–4	–10.0	–11.4
15	CCC(=O)C#C	106	24	6.8	7.0
21	CC(C)=CC=O	133	18	9.7	19.2
22	CC=C(C)C=O	117	3	10.2	11.4
32	CC=CC=CC=O	174	2	0.5	11.8
33	C#CC(=O)C(C)C	118	20	5.0	4.9
34	CC=C(C=C)C=O	144	8	18.1	11.1
35	CCC=CCC=O	121	–19	–5.8	–12.5
62	CC(=C)CCC(C)=O	150	15	9.7	8.1
88	CCC(C)=C(C)C(C)=O	158	–16	–12.4	–7.7
119	CCC(C)CC(=O)CCC	161	–18	–19.8	–17.4
122	CCC(C)C(=O)C(C)CC	162	–6	–11.5	–4.5
154	CC(C)OC(=O)C=C	110	–7	–8.1	–13.8
169	CC(C)COC(=O)C=C	132	–4	–8.8	–13.1
185	CC(C)C=CCOC(C)=O	172	–4	8.7	11.3
199	CCC(C)C(=O)OC(C)C	144	–14	–2.8	–0.4
200	CCCCCCCCOC=O	178	–12	–14.2	–11.3

[a] The structures of the carbonyl compounds are presented in Figure 7
[b] Residuals from Ref. 65 for all 200 compounds
[c] Residuals computed with Eq. (4) for aldehydes and ketones and with Eq. (5) for esters
[d] Residuals computed with Eq. (9) for all 200 compounds

For some compounds, the QSPR models obtained in this study represent a significant reduction of the residuals; in this category we find in Table 4 compound 1 (residuals –24 from Eq. (3), 5.3 from Eq. (4), and 4.3 from Eq. (9)), compound 2 (residuals –20 from Eq. (3), –9.2 from Eq. (4), and –9.9 from Eq. (9)), compound 15 (residuals 24 from Eq. (3), 6.8 from Eq. (4), and 7.0 from Eq. (9)), compound 33 (residuals 20 from Eq. (3), 5.0 from Eq. (4), and 4.9 from Eq. (9)), compound 199 (residuals –14 from Eq. (3), –2.8 from Eq. (4), and –0.4 from Eq. (9)). Together with the better statistical indices of the QSPR models presented in this study, this improvement in the computed boiling temperature of several carbonyl compounds that have large residuals with the model from Eq. (3) shows that the geometric, charge partial surface area, and quantum descriptors are essential in obtaining better correlations for polar compounds.
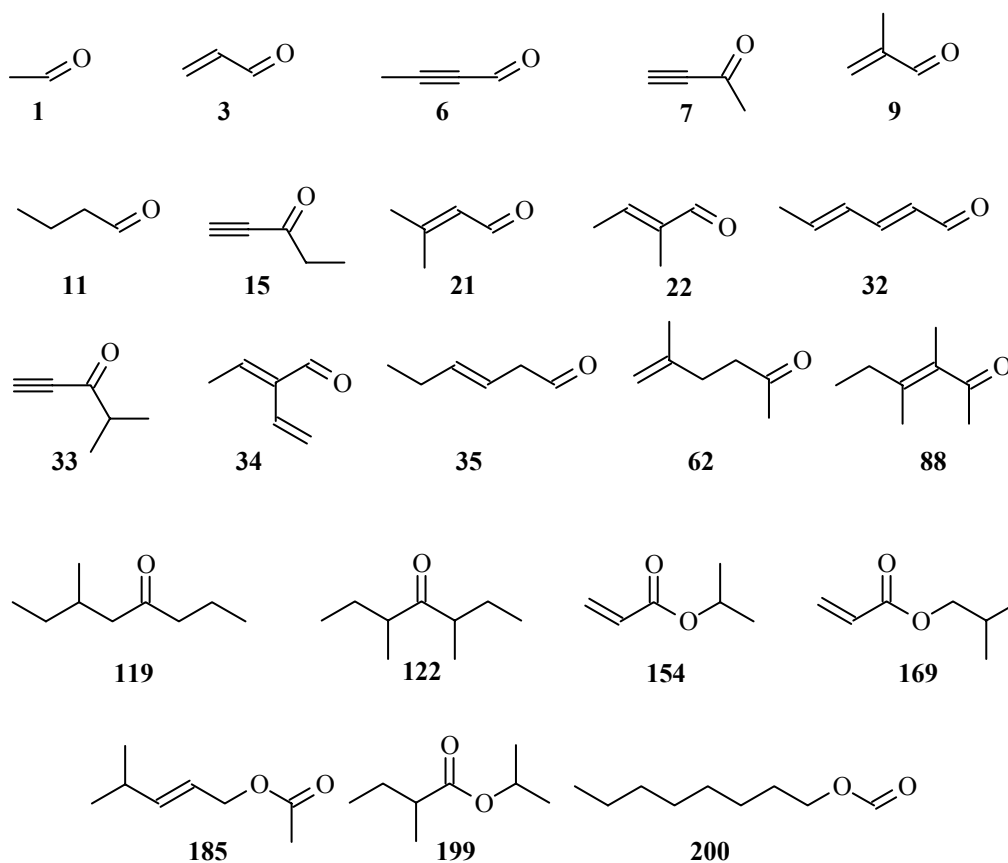
Figure 7. Strucure of the carbonyl compounds from Table 4.

However, in Table 4 we find also compounds with small residuals computed with Eq. (3) and larger errors in Eqs. (4), (5), or (9): compounds 6, 11, 22, 169, and 185. Compounds 6, 32, and 154 exhibit a different behavior: it has small residuals computed with Eqs. (3) and (4), and a large error in Eq. (9). Although the global statistical indices show that Eq. (9) is better that Eq. (3), the above results point to molecules with inferior predictions with Eq. (9); such cases emphasize the difficulties related to QSPR predictions for novel compounds. The comparative analysis of these QSPR models obtained with different groups of structural descriptors suggests an efficient way to improve the prediction of the boiling temperature, namely a parallel use of several QSPR equations obtained with different descriptors and having similar statistical indices. For a given compound, the predictions obtained with different QSPR models are averaged, and whenever a prediction deviates too much from the mean, its value is eliminated from the average. This simple procedure can detect the failure of a certain QSPR equation for a given compound, and provides a more reliable prediction than any single model.

4 CONCLUSIONS

Many QSPR models are developed for very diverse databases of chemicals, with the intention to predict a certain property for a large diversity of organic compounds. These structure–property

models allow a rapid estimation of the property for any organic compound, but the predictions are generally affected by significant errors. The present study indicates that an improvement in the prediction of the boiling temperatures of organic compounds can be obtained by developing models for classes of structurally related compounds. For a group of 200 acyclic carbonyl compounds we have modeled the boiling temperature using a large diversity of structural descriptors, i.e. constitutional, topological, geometric, electrostatic and quantum indices, that can be easily computed with standard quantum chemistry packages. Three groups of computational experiments were conducted, by considering aldehydes and ketones in one group, esters in a second group, and a third group by unifying the first two sets. Good QSPR models were obtained with five structural descriptors, with slightly lower statistics when all compounds are combined into a single set : 127 aldehydes and ketones ($r = 0.990$, $r_{LOO} = 0.986$, $s = 5.3$ °C, and $F = 1190$); 73 esters ($r = 0.993$, $r_{LOO} = 0.991$, $s = 4.2$ °C, and $F = 906$); all 200 carbonyl compounds ($r = 0.988$, $r_{LOO} = 0.987$, $s = 5.6$ °C, and $F = 1628$). In monoparametric correlations the molecular polarizability α gives the best QSPR equations, while in models with five parameters the most important descriptors depend on the separation of the compounds in different groups: the Randić connectivity index of order 1 $^1\chi$, the molecular polarizability β, and the minimum exchange energy for a C–C bond for aldehydes and ketones; the Randić connectivity index $^1\chi$, the molecular polarizability β, and the maximum atomic orbital electronic population for esters; the molecular polarizability α, the minimum valency of a carbon atom, and the Kier flexibility index for all carbonyl compounds.

5 REFERENCES

[1] O. Ivanciuc and J. Devillers, Algorithms and Software for the Computation of Topological Indices and Structure–Property Models; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 779–804.

[2] A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1–18.

[3] 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.

[4] ADAPT, P. C. Jurs, 152 Davey Lab, Chemistry Department, Penn State University, University Park, PA 16802 U.S.A., Tel: 814–865–3739, E–mail pcj@psu.edu, www http://zeus.chem.psu.edu/ADAPT.html.

[5] P. J. Hansen and P. C. Jurs, Prediction of Olefin Boiling Points from Molecular Structure, *Anal. Chem.* **1987**, *59*, 2322–2327.

[6] D. T. Stanton, P. C. Jurs, and M. G. Hicks, Computer–Assisted Prediction of Normal Boiling Points of Furans, Tetrahydrofurans, and Thiophenes, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301–310.

[7] D. T. Stanton, L. M. Egolf, P. C. Jurs, and M. G. Hicks, Computer–Assisted Prediction of Normal Boiling Points of Pyrans and Pyrroles, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306–316.

[8] L. M. Egolf and P. C. Jurs, Prediction of Boiling Points of Organic Heterocyclic Compounds Using Regression and Neural Network Techniques, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 616–625.

[9] L. M. Egolf, M. D. Wessel, and P. C. Jurs, Prediction of Boiling Points and Critical Temperatures of Industrially Important Compounds from Molecular Structure, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 947–956.

[10] M. D. Wessel and P. C. Jurs, Prediction of Normal Boiling Points of Hydrocarbons from Molecular Structure, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 68–76.

[11] M. D. Wessel and P. C. Jurs, Prediction of Normal Boiling Points for a Diverse Set of Industrially Important Organic Compounds from Molecular Structure, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 841–850.

[12] E. S. Goll and P. C. Jurs, Prediction of the Normal Boiling Points of Organic Compounds from Molecular Structures with a Computational Neural Network Model, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 974–983.

[13] 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.

[14] O. G. Mekenyan, S. H. Karabunarliev, J. M. Ivanov, and D. N. Dimitrov, A New Development of the OASIS Computer System for Modeling Molecular Properties. *Comput. Chem.* **1994**, *18*, 173–187.

[15] 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.

[16] SciQSAR, SciVision, Inc., 200 Wheeler Road, Burlington, MA 01803, U.S.A., Phone: 1–781–272–4949, Fax: 1–781–272–6868, E–mail: scivision@delphi.com, www http://www.scivision.com.

[17] 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.

[18] 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.

[19] M. Karelson, V. S. Lobanov, and A. R. Katritzky, Quantum–Chemical Descriptors in QSAR/QSPR Studies, *Chem. Rev.* **1996**, *96*, 1027–1043.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] R. Hiob and M. Karelson, Quantitative Relationship between Rate Constants of the Gas–Phase Homolysis of N–N, O–O and N–O Bonds and Molecular Descriptors, *Internet Electron. J. Mol. Des.* **2002**, *1*, 193–202, http://www.biochempress.com.

[26] S. Devotta and V. R. Pendyala, Modified Joback Group Contribution Method for Normal Boiling Point of Aliphatic Halogenated Compounds, *Ind. Eng. Chem. Res.* **1992**, *31*, 2042–2046.

[27] S. E. Stein and R. L. Brown, Estimation of Normal Boiling Points from Group Contributions, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 581–587.

[28] S. Wang, G. W. A. Milne, and G. Klopman, Graph Theory and Group Contributions in the Estimation of Boiling Points, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1242–1250.

[29] P. Simamora and S. H. Yalkowsky, Group Contribution Methods for Predicting the Melting Points and Boiling Points of Aromatic Compounds, *Ind. Eng. Chem. Res.* **1994**, *33*, 1405–1409.

[30] H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.* **1947**, *69*, 17–20.

[31] J. R. Platt, Influence of Neighbor Bonds on Aditive Bond Properties in Paraffins, *J. Chem. Phys.* **1947**, *15*, 419–420; J. R. Platt, Prediction of Isomeric Differences in Paraffin Properties, *J. Phys. Chem.* **1952**, *56*, 328–336.

[32] P. A. Filip, T.–S. Balaban, and A. T. Balaban, A New Approach for Devising Local Graph Invariants: Derived Topological Indices with Low Degeneracy and Good Correlational Ability, *J. Math. Chem.* **1987**, *1*, 61–83.

[33] D. E. Needham, I.–C. Wei, and P. G. Seybold, Molecular Modeling of the Physical Properties of the Alkanes, *J. Am. Chem. Soc.* **1988**, *110*, 4186–4194.

[34] A. T. Balaban and T.–S. Balaban, New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances, *J. Math. Chem.* **1991**, *8*, 383–397.

[35] S. C. Basak and G. D. Grunwald, A Comparative Study of Graph Invariants, Total Surface Area and Volume in Predicting Boiling Points of Alkanes, *Math. Model. Sci. Comput.* **1993**, *2*, 735–740.

[36] O. Ivanciuc, T.–S. Balaban, and A. T. Balaban, Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices, *J. Math. Chem.* **1993**, *12*, 309–318.

[37] M. Randić and N. Trinajstić, Isomeric Variations in Alkanes: Boiling Points of Nonanes, *New J. Chem.* **1994**, *18*, 179–189.

[38] D. Cherqaoui and D. Villemin Use of a Neural Network to Determine the Boiling Point of Alkanes, *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 97–102.

[39] A. A. Gakh, E. G. Gakh, B. G. Sumpter, and D. W. Noid, Neural Network–Graph Theory Approach to the Prediction of the Physical Properties of Organic Compounds, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 832–839.

[40] M. V. Diudea, O. Ivanciuc, S. Nikolić, and N. Trinajstić, Matrices of Reciprocal Distance, Polynomials and Derived Numbers, *MATCH* (*Commun. Math. Comput. Chem.*) **1997**, *35*, 41–64.

[41] O. Ivanciuc, M. V. Diudea, and P. V. Khadikar, New Topological Matrices and Their Polynomials, *Ind. J. Chem.* **1998**, *37A*, 574–585.

[42] O. Ivanciuc, Artificial Neural Networks Applications. Part 9. MolNet Prediction of Alkane Boiling Points, *Rev. Roum. Chim.* **1998**, *43*, 885–894.

[43] S. Liu, C. Cao, and Z. Li, Approach to Estimation and Prediction for Normal Boiling Point (NBP) of Alkanes Based on a Novel Molecular Distance–Edge (MDE) Vector, λ, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 387–394.

[44] E. Estrada, O. Ivanciuc, I. Gutman, A. Gutierrez, and L. Rodríguez, Extended Wiener Indices. A New Set of Descriptors for Quantitative Structure–Property Studies, *New J. Chem.* **1998**, *22*, 819–822.

[45] C. Cao, S. Liu, and Z. Li, On Molecular Polarizability: 2. Relationship to the Boiling Point of Alkanes and Alcohols, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1105–1111.

[46] G. Rücker and C. Rücker, On Topological Indices, Boiling Points, and Cycloalkanes, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788–802.

[47] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, The Complementary Distance Matrix, a New Molecular Graph Metric, *A C H – Model. Chem.* **2000**, *137*, 57–82.

[48] O. Ivanciuc, T. Ivanciuc, D. Cabrol–Bass, and A. T. Balaban, Evaluation in Quantitative Structure–Property Relationship Models of Structural Descriptors Derived from Information–Theory Operators, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 631–643.

[49] G. Espinosa, D. Yaffe, Y. Cohen, A. Arenas, and F. Giralt, Neural Network Based Quantitative Structural Property Relations (QSPRs) for Predicting Boiling Points of Aliphatic Hydrocarbons, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 859–879.

[50] M. Randić, Quantitative Structure–Property Relationship. Boiling Points of Planar Benzenoids, *New J. Chem.* **1996**, *20*, 1001–1009.

[51] C. M. White, Prediction of the Boiling Point, Heat of Vaporization, and Vapor Pressure at Various Temperatures for Polycyclic Aromatic Hydrocarbons, *J. Chem. Eng. Data* **1986**, *31*, 198–203.

[52] D. Plavšić, N. Trinajstić, D. Amić, and M. Šoškić, Comparison between the Structure–Boiling Point Relationships with Different Descriptors for Condensed Benzenoids, *New J. Chem.* **1998**, *22*, 1075–1078.

[53] A. T. Balaban, N. Joshi, L. B. Kier, and L. H. Hall, Correlations between Chemical Structure and Normal Boiling Points of Halogenated Alkanes C_1–C_4, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 233–237.

[54] A. T. Balaban, S. C. Basak, T. Colburn, and G. D. Grunwald, Correlation between Structure and Normal Boiling Points of Haloalkanes C_1–C_4 Using Neural Networks, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118–1121.

[55] S. C. Basak, B. D. Gute, and G. D. Grunwald, Estimation of the Normal Boiling Points of Haloalkanes Using Molecular Similarity, *Croat. Chem. Acta* **1996**, *69*, 1159–1173.

[56] T. S. Carlton, Correlation of Boiling Points with Molecular Structure for Chlorofluoroethanes, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 158–164.

[57] J. Wei, Boiling Points and Melting Points of Chlorofluorocarbons, *Ind. Eng. Chem. Res.* **2000**, *39*, 3116–3119.

[58] A. T. Balaban, L. B. Kier, and N. Joshi, Correlations between Chemical Structure and Normal Boiling Points of Acyclic Ethers, Peroxides, Acetals, and Their Sulfur Analogues, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 237–244.

[59] H. Lohninger, Evaluation of Neural Networks Based on Radial Basis Functions and Their Application to the Prediction of Boiling Points from Structural Parameters, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 736–744.

[60] D. Cherqaoui, D. Villemin, A. Mesbah, J.–M. Cense, and V. Kvasnicka, Use of a Neural Network to Determine the Boiling Points of Acyclic Ethers, Peroxides, Acetals and their Sulfur Analogues, *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 2015–2019.

[61] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban Design of Topological Indices. Part 10. Parameters Based on Electronegativity and Covalent Radius for the Computation of Molecular Graph Descriptors for Heteroatom–Containing Molecules, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 395–401.

[62] M. Randić and S. C. Basak, Construction of High–Quality Structure–Property–Activity Regressions: The Boiling Points of Sulfides, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 899–905.

[63] L. H. Hall and C. T. Story, Boiling Point of a Set of Alkanes, Alcohols and Chloroalkanes: QSAR with Atom Type Electrotopological State Indices Using Artificial Neural Networks, *SAR QSAR Environ. Res.* **1997**, *6*, 139–161.

[64] A. P. Bünz, B. Braun, and R. Janowsky, Application of Quantitative Structure–Performance Relationship and Neural Network Models for the Prediction of Physical Properties from Molecular Structure, *Ind. Eng. Chem. Res.* **1998**, *37*, 3043–3051.

[65] A. T. Balaban, D. Mills, and S. C. Basak, Correlation between Structure and Normal Boiling Points of Acyclic Carbonyl Compounds, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 758–764.

[66] A. T. Balaban, S. C. Basak, and D. Mills, Normal Boiling Points of 1,ω–Alkanedinitriles: The Highest Increment in a Homologous Series, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 769–774.

[67] D. T. Stanton, Development of a Quantitative Structure–Property Relationship Model for Estimating Normal Boiling Points of Small Multifunctional Organic Molecules, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 81–90.

[68] L. H. Hall and C. T. Story, Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electrotopological State Indices Using Artificial Neural Networks, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004–1014.

[69] R. Abramowitz and S. H. Yalkowsky, Melting Point, Boiling Point and Symmetry, *Pharm. Res.* **1990**, *7*, 942–947.

[70] R. L. Rich, Boiling Point and the Refraction (Polarizability) of Exposed Atoms, *Bull. Chem. Soc. Japan* **1993**, *66*, 1065–1078.

[71] P. Simamora and S. H. Yalkowsky, Quantitative Structure Property Relationship in the Prediction of Melting Point and Boiling Point of Rigid Non–Hydrogen Bonding Organic Molecules, *SAR QSAR Environ. Res.* **1993**, *1*, 293–300.

[72] J. S. Murray, P. Lane, T. Brinck, K. Paulsen, M. E. Grice, and P. Politzer, Relationships of Critical Constants and Boiling Points to Computed Molecular Surface Properties, *J. Phys. Chem.* **1993**, *97*, 9369–9373.

[73] P. Simamora, A. H. Miller, and S. H. Yalkowsky, Melting Point and Normal Boiling Point Correlations: Applications to Rigid Aromatic Compounds, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 437–440.

[74] S. H. Yalkowsky, J. F. Krzyaniak, and P. B. Myrdal, Relationships between Melting Point and Boiling Point of Organic Compounds, *Ind. Eng. Chem. Res.* **1994**, *33*, 1872–1877.

[75] 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.

[76] I. N. Tsibanogiannis, N. S. Kalospiros, and D. P. Tassios, Prediction of Normal Boiling Point Temperature of Medium/High Molecular Weight Compounds, *Ind. Eng. Chem. Res.* **1995**, *34*, 997–1002.

[77] J. F. Krzyaniak, P. B. Myrdal, P. Simamora, and S. H. Yalkowsky, Boiling Point and Melting Point Prediction for Aliphatic, Non–Hydrogen–Bonding Compounds, *Ind. Eng. Chem. Res.* **1995**, *34*, 2530–2535.

[78] S. C. Basak, B. D. Gute, and G. D. Grunwald, A Comparative Study of Topological and Geometrical Parameters in Estimating Normal Boiling Point and Octanol/Water Partition Coefficient, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1054–1060.

[79] J. Tetteh, T. Suzuki, E. Metcalfe, and S. Howells, Quantitative Structure–Property Relationships for the Estimation of Boiling Point and Flash Point Using a Radial Basis Function Neural Network, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 491–507.

[80] A. T. Balaban, Highly Discriminating Distance–Based Topological Index, *Chem. Phys. Lett.* **1982**, *89*, 399–404.

[81] A. T. Balaban, Topological Indices Based on Topological Distances in Molecular Graphs, *Pure Appl. Chem.* **1983**, *55*, 199–206.

[82] A. T. Balaban, Chemical Graphs. Part 48. Topological Index J for Heteroatom–Containing Molecules Taking Into Account Periodicities of Element Properties, *MATCH* (*Commun. Math. Chem.*) **1986**, *21*, 115–122.

[83] A. T. Balaban and O. Ivanciuc, FORTRAN 77 Computer Program for Calculating the Topological Index J for Molecules Containing Heteroatoms; in: *MATH/CHEM/COMP 1988*, Ed. A. Graovac, Studies in Physical and Theoretical Chemistry; Elsevier: Amsterdam, 1989, Vol. 63, pp 193–212.

[84] S. C. Basak, Information Theoretic Indices of Neighborhood Complexity and Their Applications; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 563–593.

[85] HyperChem 5.1, Hypercube, Inc., Florida Science and Technology Park, 1115 N.W. 4th Street Gainesville, Florida 32601, U.S.A., E–mail info@hyper.com, www http://www.hyper.com.

[86] MOPAC 6, adapted for Windows by V. Lobanov, www http://www.ccl.net.

[87] 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.

[88] CODESSA 2.13, Semichem, 7204 Mullen, Shawnee, KS 66216, U.S.A., E–mail andy@semichem.com, www http://www.semichem.com.

[89] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons, New York, 2000.

[90] D. T. Stanton and P. C. Jurs, Development and Use of Charged Partial Surface Area Structural Descriptors in Computer–Assisted Quantitative Structure–Property Relationship Studies. *Anal. Chem.* **1990**, *62*, 2323–2329.