Pergamon

Tetrahedron 54 (1998) 9129–9142

TETRAHEDRON

Quantitative Structure-Property Relationship Study of Normal Boiling Points for Halogen-/ Oxygen-/ Sulfur-Containing Organic Compounds Using the CODESSA Program*

Ovidiu Ivanciuc, Teodora Ivanciuc and Alexandru T. Balaban

University "Politehnica" of Bucharest, Faculty of Chemical Technology
Department of Organic Chemistry,
PO Box 12 - 243, 78100 Bucharest, Romania
E-mail: o_ivanciuc@chim.upb.ro; balaban@acad.ro

Received 20 February 1998; revised 22 April 1998; accepted 3 June 1998

***Abstract*:** QSPR (Quantitative Structure-Property Relationship) models for the estimation of boiling points of organic compounds containing halogens, oxygen, or sulfur without hydrogen bonding were established with the CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) program developed by Katritzky and coworkers. The boiling points of 185 compounds containing oxygen or sulfur can be accurately computed with a MLR (Multi-Linear Regression) equation containing six theoretical descriptors and having the following statistical indices: $r = 0.992$ and $s = 6.3$ ^{o}C. For a set of 534 halogenated alkanes C_1 – C_4 the best MLR equation with five descriptors has $r = 0.990$ and $s = 9.0$ ^{o}C. For subsets of molecules corresponding to monoethers, monosulfides, F- and Cl-containing compounds or F-containing compounds, even better MLR models were obtained, showing that a greater accuracy can be obtained for the estimation of boiling points with QSPR models developed for restricted classes of compounds. The QSPR models developed with CODESSA allow accurate computation of the boiling points of organic compounds using simple constitutional, topological, electrostatic and quantum indices that can be computed with standard quantum chemistry packages.

INTRODUCTION

The boiling point of a compound is an important property for the simulation of processes in chemical and petroleum industries. With the increased need of reliable data for optimization of industrial processes, it is important to develop QSPR (Quantitative Structure-Property Relationship) models for the estimation of normal boiling point for compounds that are not yet synthesized or whose boiling point is unknown.[1-3] The usual group additivity model is no longer the best choice in developing structure-property models due to the high precision required and to the high diversity of chemical compounds containing multiple functionalities.

General strategies for generating structure-property models were proposed along the years. The ADAPT program[4] is designed to compute the 3-dimensional structure of the investigated chemical comopunds, 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 boiling points for organic compounds.[5]

OASIS[6] (Optimized Approach based on Structural Indices Selection) has mainly the same functions, i.e. it computes a large set of descriptors and then uses them in various linear and nonlinear models. This program was applied in devising structure-activity relationships.[7]

PII: S0040-4020(98)00550-X

The CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) program developed by Katritzky and coworkers represents another QSPR approach which computes about 500 structural parameters using constitutional, topological, geometrical, electrostatic and quantum descriptors of the chemical compounds.[8] The computed structural descriptors are used to develop models of the investigated property. CODESSA was successfully employed in QSPR studies concerning the prediction of gas chromatographic retention indices, solubility of gases and vapors in water, polymer glass transition temperatures, critical micelle concentration, melting points, and boiling points.[9-11] The main advantage of CODESSA over other statistical packages used in developing QSPR models is the easy generation of a large number of theoretical descriptors coding in numerical form the chemical structure.

Numerous QSPR models for calculating the boiling points of organic compounds have been introduced using various numerical descriptors of chemical structure. Empirical models were proposed,[12] in which the boiling point was computed on the basis of a set of known molecular properties. Such a QPPR (Quantitative Property-Property Relationship) model requires the experimental determination of properties used as independent variables; if such experimental values are missing, then the model is of little use.

Molecular group contribution methods are widely employed to estimate boiling points.[13] The difficulty of this approach is represented by the definiton 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.

The largest number of QSPR models for boiling point estimation were developed by describing the chemical structure with the aid of various topological indices.[14-19] Topological indices are numerical representations of the chemical structure computed on the basis of the molecular graph. The computation of topological indices is very simple when compared to the computation of geometric or quantum descriptors.

Quantum descriptors were also proposed for the computation of boiling points, particularly those related to molecular surfaces.[20]

Neural networks, representing general nonlinear models, were also used with considerable success in developing QSPR models for boiling points. Various neural structures were used, such as the multilayer feed-forward,[21-23] radial basis function,[24] or MolNet.[25]

Particularly important is the estimation of boiling points of halogen-containing organic compounds. Chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs) are being used as solvents, foaming agents and refrigerants, while bromochlorofluorocarbons (BCFCs) are used as fire extinguishing agents. The use of CFCs has damaging effects on the stratospheric ozone layer; all halogenated compounds have an important contribution in global warming by the green house effect. The development of alternatives to CFCs, compounds with zero or low ozone depletion potential and a low global warming potential, has generated a high interest in predicting boiling points of CFCs and HCFCs based only on molecular structure. Such a screening can use the boiling point to compute the volatility and the required thermodynamic properties, thus directing the synthetic efforts only to the promising candidates.

The purpose of this study is to develop QSPR models for the estimation of boiling points of organic compounds containing halogens, oxygen, or sulfur using the CODESSA program.

QSPR METHOD

Data and Previous Results

The first set of molecules used in the QSPR investigation of normal boiling point, denoted the **O-S** set, contains 185 saturated acyclic compounds possessing divalent oxygen or sulfur atoms, and free from hydrogen bonding, from the following classes: 73 ethers, 17 diethers, 21 acetals, 6 peroxides, 45 sulfides (thioethers), 6-bis-sulfides, 4 thioacetals, and 13 disulfides. The experimental boiling points for the **O-S** set were collected from Beilstein's *Handbuch der Organischen Chemie*, from the CRC *Handbook of Physics and Chemistry*,[26] and the *Dictionary of Organic Compounds*.[27] Previously, this **O-S** set was used in a QSPR study that used constitutional and molecular graph descriptors.[16] For the whole set, a two-variable equation gave satisfactory results:

$$\mathbf{BP} = -59.10(\pm3.11) + 44.30(\pm0.76)\,{}^1\chi + 42.88(\pm0.95)\mathbf{NoS} \qquad \text{(A)}$$

$$n = 185 \qquad r = 0.982 \qquad s = 9.0\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 2390$$

where ${}^1\chi$ is the Randić connectivity index, and **NoS** is the number of sulfur atoms. A good three-variable equation was obtained by considering the electrotopological state of the heteroatoms $\mathbf{S}_{\mathrm{het}}$, the Kier and Hall valence connectivity index ${}^0\chi^v$, and the number of methyl groups **NoMe**:

$$\mathrm{BP} = -11.23 - 7.21\mathbf{S}_{\mathrm{het}} + 35.04\,{}^0\chi^v - 18.30\mathbf{NoMe} \qquad \text{(B)}$$

$$n = 185 \qquad r = 0.977 \qquad s = 10.5\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 1019$$

The use of the topological index **J** modified for the presence of heteroatoms, $\mathbf{J}_{\mathrm{het}}$, improved the QSPR equation:

$$\mathrm{BP} = -41.75(\pm3.90) + 43.79(\pm0.69)\,{}^1\chi + 45.03(\pm0.90)\mathbf{NoS} - 2.90(\pm0.41)\mathbf{J}_{\mathrm{het}} \qquad \text{(C)}$$

$$n = 185 \qquad r = 0.986 \qquad s = 8.2\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 2048$$

Two subsets were formed from the **O-S** set, namely a subset of 72 monoethers and a subset of 44 monosulfides. The first term in each of the two series (Me_2O and Me_2S, respectively) were discarded. For the set of monoethers two equations were developed:

$$\mathrm{BP} = -51.08(\pm9.13) + 46.99(\pm0.95)\,{}^1\chi - 3.93(\pm1.45)\mathbf{J}_{\mathrm{het}} \qquad \text{(D)}$$

$$n = 72 \qquad r = 0.989 \qquad s = 6.4\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 1501$$

$$\mathrm{BP} = 46.45(\pm16.02) - 25.86(\pm3.81)\mathrm{S_O} - 12.50(\pm0.66)\mathbf{NoMe} + 37.89(\pm0.70)\,{}^0\chi^v \qquad \text{(E)}$$

$$n = 72 \qquad r = 0.994 \qquad s = 4.9\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 1749$$

where $\mathrm{S_O}$ represents the electrotopological state for oxygen. The boiling points of the 44 sulfides were modeled by the following equations:

$$\mathrm{BP} = 20.01(\pm7.49) + 43.93(\pm0.83)\,{}^1\chi - 6.34(\pm1.20)\mathbf{J}_{\mathrm{het}} \qquad \text{(F)}$$

$$n = 44 \qquad r = 0.994 \qquad s = 4.3\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 1807$$

$$\mathrm{BP} = 50.59(\pm23.95) - 32.91(\pm13.45)\mathrm{S_S} - 9.86(\pm0.68)\mathbf{NoMe} + 24.10(\pm0.54)\,{}^1\kappa_\alpha \qquad \text{(G)}$$

$$n = 44 \qquad r = 0.994 \qquad s = 4.3\ {}^\circ\mathrm{C}$$

where $\mathrm{S_S}$ represents the electrotopological state for sulfur and ${}^1\kappa_\alpha$ is the Kier shape index.

Recently we have proposed a general scheme for computing parameters in vertex- and edge-weighted molecular graphs.[17] Using the molecular graph descriptors computed with the weighting schemes, we have derived a set of QSPR models for the **O-S** set from which we present two equations, one with four variables and the second with five variables:

$$\mathrm{BP} = -102.61 + 28.00\mathbf{NoO} + 141.86\mathbf{NoS} + 49.35\mathbf{Wi(A,Z)} - 9.40\mathbf{Wi(RD,Z)} \qquad \text{(H)}$$

$$n = 185 \qquad r = 0.989 \qquad s = 7.4\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 1933$$

$$\begin{aligned}\mathrm{BP} = {}& -104.02 + 52.94\mathbf{NoS} - 15.38\,{}^2\chi^v - 3.05\mathbf{MinSp(D,X)} + 75.33\mathbf{MaxSp(RD,Y)} - \\ & -15.28\mathbf{IB(D,X)}\end{aligned} \qquad \text{(I)}$$

$$n = 185 \qquad r = 0.990 \qquad s = 6.9\ {}^\circ\mathrm{C} \qquad \mathrm{F} = 1741$$

where **NoO** is the number of oxygen atoms, ${}^2\chi^v$ is the Kier and Hall valence connectivity index, **Wi** is the Wiener operator, **MinSp** and **MaxSp** are the minimum and maximum eigenvalue operator, respectively, and **IB** is the Ivanciuc-Balaban operator, **A** is the adjacency matrix, **RD** is the reciprocal distance matrix, and **X**, **Y**, and **Z** represent weighting schemes for the molecular graphs. The above results show that the molecular graph descriptors are well suited for the estimation of the boiling points of the **O-S** set of compounds.

The boiling points of the compounds in the **O-S** set were also modeled with a multi-layer feed-forward neural network, using as input the counts for 20 clusters (subgraphs) containing C, O, or S atoms.[23] The best leave-one-out cross-validation results were obtained with a network containing 5 hidden neurons, and therefore we present and comment here only the results obtained with this type of network. The calibration correlation coefficient and standard deviation, $r_{\mathrm{cal}} = 0.998$ and $s_{\mathrm{cal}} = 2.9$ °C, and the cross-validation statistical indices, $r_{\mathrm{cv}} = 0.989$ and $s_{\mathrm{cv}} = 5.1$ °C, are superior to those obtained in MLR models, but this neural network has 111 adjustable parameters (connections between neurons). The linear models presented above have only 3 or 4 parameters, and therefore we consider that the improvement in estimating the boiling point for the **O-S** set is obtained with a too large number of parameters.

Another neural network study of the **O-S** set used radial basis functions networks with three input neurons, 20 hidden neurons, and one output neuron.[24] Using as input data ${}^1\chi$, **NoS**, and $\mathbf{J}_{\mathrm{het}}$ the neural model offered better statistical indices than the multilinear one, i.e. $r = 0.993$ and $s = 5.8$ °C, at the expense of a much larger number of adjustable parameters.

The second set of molecules used in the QSPR investigation of normal boiling points, denoted the **Hal** set, contains 534 halo- and polyhaloalkanes with 1-4 carbon atoms. The experimental boiling points for the **Hal** set were collected from Beilstein's Handbuch der Organischen Chemie, from the CRC Handbook of Physics and Chemistry,[26] from the Dictionary of Organic Compounds,[27] and from ref. 28. The **Hal** set was used in a previous QSPR study that considered a MLR model with constitutional and molecular graph descriptors.[18] To the set of 532 compounds used in ref. 18 we have added 1,1,1,2-tetrachloro-2,2-difluoroethane with the boiling point 92.6 °C and 1,1,2-trifluoro-1-iodoethane with 101 °C. For the whole set, several equations were reported:

$$\begin{aligned} \mathrm{BP} = &- 73.65(\pm 1.98) + 33.21(\pm 2.21)[{}^1\chi^v - {}^0\chi^v] - 64.06(\pm 0.79)\,{}^D\chi^0 + 98.46(\pm 2.11)\,{}^1\chi - \\ &- 20.65(\pm 0.66)\mathbf{NoBr} - 22.18(\pm 1.30)\mathbf{NoI} + 6.36(\pm 0.83)[{}^2\chi^v - {}^1\chi^v] \end{aligned} \tag{J}$$

$n = 532$ $r = 0.985$ $s = 10.94$ °C F = 2953

$$\begin{aligned} \mathrm{BP} = &- 81.57(\pm 2.03) + 6.29(\pm 1.12)\phi - 53.46(\pm 1.50)\,{}^D\chi^0 + 71.76(\pm 3.10)\,{}^1\chi - \\ &- 19.43(\pm 0.70)\mathbf{NoBr} - 22.15(\pm 1.33)\mathbf{NoI} + 10.03(\pm 2.16)[{}^1\chi^v - {}^0\chi^v] \end{aligned} \tag{K}$$

$n = 532$ $r = 0.985$ $s = 11.20$ °C F = 2814

$$\begin{aligned} \mathrm{BP} = &- 53.78(\pm 3.27) + 138.62(\pm 5.63)\,{}^1\chi - 15.35(\pm 1.33)\mathbf{NoH} - 76.35(\pm 3.45)\mathbf{NoF} - \\ &- 37.21(\pm 3.48)\mathbf{NoCl} - 14.29(\pm 3.54)\mathbf{NoBr} + 14.45(\pm 3.73)\mathbf{NoI} \end{aligned} \tag{L}$$

$n = 532$ $r = 0.985$ $s = 11.59$ °C F = 2620

where **NoH**, **NoF**, **NoCl**, **NoBr**, **NoI** are the numbers of H, F, Cl, Br, and I atoms, respectively, ϕ is the Kier flexibility index, and ${}^D\chi^0$ is the Kier and Hall difference connectivity index.

A subset of the **Hal** set of compounds was used in a QSPR study that estimated the boiling points with a multi-layer feed-forward neural network, using as input the counts for C, F, Cl, Br, and I atoms and the Wiener **W** and Balaban **J** topological indices.[22a] For a subset of 276 haloalkanes containing only F and/or Cl, a neural network model with 5 input and 10 hidden neurons, with 61 adjustable parameters, offered good statistical indices, namely $r = 0.992$ and $s = 8.5$ °C.

A set of 267 chlorofluorocarbons (CFCs) from the **Hal** set was used in a similarity QSPR model.[19] The structure of each compound was described with the aid of 59 topological indices, mainly connectivity and information indices, and the Wiener **W** and Balaban **J** indices. The boiling point of a certain CFC is equal to the mean boiling point of the K-nearest neighbours in the molecular structure space with 59 dimensions determined by the set of topological indices. For K = 5, one could arrive at $r = 0.933$ and $s = 23.7$ °C, showing that the similarity QSPR model can extract the information from a space with a high dimension.

Structural Descriptors

The structures were drawn with HyperChem[29] and exported in a file format suitable for AMPAC. The geometry optimization was performed on a Pentium 133 MHz with the semiempirical quantum method AM1[30] using the AMPAC 5.0 program.[31] The HyperChem structure files and the AMPAC output files were used by the CODESSA program[32] to compute more than 500 structural descriptors. 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 shape indices, etc.); geometrical (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, LUMO, etc.).

Multilinear Regression Model

From the whole set of 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. CODESSA was used to develop multilinear regression QSAR models by a heuristic method which included the following steps:

(a) all orthogonal pairs of structural descriptors were selected from the initial set. Two descriptors are considered orthogonal if the 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 were used in the third step.

(c) to a MLR model containig 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.

RESULTS

Correlations for the Normal Boiling Points of Oxygen- and Sulfur-Containing Compounds Devoid of Hydrogen Bonding (O-S Set)

Using the HyperChem structural files and AMPAC AM1 output files for the 185 compounds in the **O-S** set CODESSA generated 391 theoretical descriptors. Table 1 presents the notation of the 63 descriptors involved in the QSPR correlations for the **O-S** set. The descriptor type is coded with C for constitutional, E for electrostatic, G for geometric, Q for quantum, and T for topological. The definitions of the descriptors are presented in the CODESSA Reference and User's Manuals. The theoretical descriptors computed with CODESSA were used in monoparametric correlations relating the experimental normal boiling point to the structural indices, and those offering the best results were selected for further use. The number of descriptors was reduced to 283 by eliminating those with F-test values less than 1, t-test values less than 0.1 or correlation coefficients with the boiling point less than 0.1.

Descriptors will be coded according to Tables 1 and 3 by their numbers in square brackets, and correlation equations by boldface numbers.

The best five mono-parametric correlations are collected in Table 2, Eqs. (**1-5**), giving the descriptor involved, the correlation coefficient r, the standard deviation s and the F-test value. From the 5 descriptors, 2 are topological, 2 are quantum, and one (the molecular weight) is constitutional. The Kier and Hall ${}^{1}\chi^{v}$ index, with $r = 0.961$ and $s = 13.4$ °C, is by far the best descriptor in mono-parametric equations, surpassing the more complex electrostatic or quantum descriptors.

Biparametric correlations were computed by using all pairs of descriptors with an intercorrelation coefficient lower than 0.8. The best five biparametric correlations are presented in Table 2, Eqs. (**6-10**). The introduction of a second parameter significantly improves the statistical indices of the QSPR model, the standard deviation decreasing below 10 °C. The best biparametric correlation, Eq. (**6**), has the same descriptors as in Eq. (A), with minor differences in the statistical indices due to the different programs used to develop the

Table 1. Notation of the Descriptors Involved in the QSPR Equations from Table 2 that Compute the Boiling Points of Oxygen and Sulfur Containing Compounds

No.	Type	Descriptors
[1]	C	Molecular weight
[2]	C	Number of S atoms
[3]	C	Relative number of S atoms
[4]	E	FPSA-2 Fractional PPSA (PPSA-2/TMSA) (electrostatic)
[5]	E	Maximum partial charge for a C atom (electrostatic)
[6]	E	FPSA-3 Fractional PPSA (PPSA-3/TMSA) (electrostatic)
[7]	E	PPSA-3 Atomic charge weighted PPSA (electrostatic)
[8]	E	PNSA-3 Atomic charge weighted PNSA (electrostatic)
[9]	E	WNSA-3 Weighted PNSA (PNSA3*TMSA/1000) (electrostatic
[10]	E	WNSA-1 Weighted PNSA (PNSA1*TMSA/1000) (electrostatic)
[11]	E	FNSA-2 Fractional PNSA (PNSA-2/TMSA) (electrostatic)
[12]	E	RPCG Relative positive charge (QMPOS/QTPLUS) (electrostatic)
[13]	E	WNSA-2 Weighted PNSA (PNSA2*TMSA/1000) (electrostatic)
[14]	E	RNCS Relative negative charged SA (SAMNEG*RNCG) (electrostatic)
[15]	E	FPSA-2 Fractional PPSA (PPSA-2/TMSA) (electrostatic)
[16]	E	Topographic electronic index (all bonds) (electrostatic)
[17]	E	Maximum partial charge (Qmax)
[18]	E	PNSA-1 Partial negative surface area (electrostatic)
[19]	E	Minimum partial charge for a C atom (electrostatic)
[20]	G	Molecular volume
[21]	G	Molecular surface area
[22]	G	Moment of inertia C
[23]	Q	FPSA-2 Fractional PPSA (PPSA-2/TMSA) (quantum)
[24]	Q	Final heat of formation
[25]	Q	α polarizability (computed from the dipole moment)
[26]	Q	PPSA-3 Atomic charge weighted PPSA (quantum)
[27]	Q	Exchange energy + electron-electron repulsion for a C-O bond
[28]	Q	Maximum total interaction for a C-O bond
[29]	Q	LUMO+1 energy
[30]	Q	Maximum electron-nucleus attraction for a C-O bond
[31]	Q	Maximum resonance energy for a C-H bond
[32]	Q	Minimum electrophilic reactivity index for a C atom
[33]	Q	Minimum nucleophilic reactivity index for a C atom
[34]	Q	Minimum valency of a S atom
[35]	Q	Minimum electrophilic reactivity index for a S atom
[36]	Q	Minimum (>0.1) bond order of a H atom
[37]	Q	Minimum total interaction for a C-H bond
[38]	Q	Minimum electron-nucleus attraction for a C-H bond
[39]	Q	HOMO-1 energy
[40]	Q	^{1}IC, average information content (order 1)
[41]	Q	Maximum electrophilic reactivity index for a C atom
[42]	Q	XY Shadow / XY Rectangle
[43]	Q	Minimum electron-nucleus attraction for a H atom
[44]	Q	Minimum electron-nucleus attraction for a C-H bond
[45]	Q	Maximum nucleus-nucleus repulsion for a C-H bond
[46]	Q	Maximum electron-nucleus attraction for a C-H bond
[47]	Q	WNSA-3 Weighted PNSA (PNSA3*TMSA/1000) (quantum)
[48]	Q	Minimum resonance energy for a C-H bond
[49]	Q	FNSA-3 Fractional PNSA (PNSA-3/TMSA) (quantum)
[50]	Q	Average valency of a H atom
[51]	Q	Average bond order of a H atom
[52]	Q	WNSA-2 Weighted PNSA (PNSA2*TMSA/1000) (quantum)
[53]	Q	β polarizability (computed from the dipole moment)
[54]	Q	Minimum nucleus-nucleus repulsion for a S-S bond
[55]	Q	WPSA-2 Weighted PPSA (PPSA2*TMSA/1000) (quantum)
[56]	Q	Minimum (>0.1) bond order of a C atom
[57]	Q	Minimum electrophilic reactivity index for a O atom
[58]	T	$^1\chi$, Randić index (order 1)
[59]	T	$^1\chi^v$, Kier and Hall index (order 1)
[60]	T	$^0\chi^v$, Kier and Hall index (order 0)
[61]	T	$^3\kappa$, Kier shape index (order 3)
[62]	T	$^2\chi$, Randić index (order 2)
[63]	T	$^2\chi^v$, Kier and Hall index (order 2)

two equations. All 5 biparametric correlations contain a topologic index, either ${}^1\chi$ or ${}^1\chi^v$; the second descriptor is either **NoS** as in Eq. (**6**), or a charge-surface area descriptor.

Using the set of bi-parametric correlations obtained in the previous step, a heuristic method was applied to generate correlations with up to 6 parameters. The best five such correlations with 4, 5, and 6 parameters are presented in Table 2. An inspection of the Eqs. (**11-25**) from Table 2 shows that one topological index, either ${}^1\chi$ or ${}^1\chi^v$, is present in every equation, while the largest number of descriptors from a single class is formed by the electrostatic charge-surface area descriptors. All best 5 correlations with 6 descriptors use descriptors [59], ${}^1\chi^v$, the charge-surface area descriptors [8], WNSA-3, and [13], WNSA-2, the relative positive charge [12], RPCG, and the maximum partial charge [17], Q_{max}. The best 6 parameters correlation, Eq. (**21**), is presented below:

$$\begin{gathered} BP = 95.33(\pm 15.14) + 41.82(\pm 0.96)[59] - 19.78(\pm 1.50)[9] - 501.88(\pm 32.46)[12] + \\ + 5.57(\pm 0.48)[13] + 755.09(\pm 88.30)[17] + 8.58(\pm 1.02)[39] \\ n = 185 \qquad r = 0.992 \qquad s = 6.3\ {}^{\circ}C \qquad F = 1764.7 \end{gathered} \tag{21}$$

The correlations for the subset of 73 monoethers are presented in Table 2, Eqs. (**26-75**). The best indices in mono-parametric correlations are the Randić ${}^1\chi$ index, the fractional positive surface area FPSA-2, the Kier and Hall ${}^1\chi^v$ index, the molecular volume and the molecular surface area. The best equation with 4 variables, Eq. (**36**), is presented:

$$\begin{gathered} BP = -2191.5(\pm 197.63) + 46.34(\pm 0.78)[58] + 7.21(\pm 0.84)[27] - \\ - 134.55(\pm 18.96)[22] + 701.84(\pm 116.39)[36] \\ n = 73 \qquad r = 0.996 \qquad s = 4.0\ {}^{\circ}C \qquad F = 2227.0 \end{gathered} \tag{36}$$

Another set of correlations was developed with for the 45 monosulfides. The results presented in Table 2, Eqs. (**51-75**), show again that the best descriptors are ${}^1\chi^v$ and ${}^1\chi$, folowed by the FPSA-2 computed with quantum charges, the heat of formation, and the molecular surface area. A correlation with 4 descriptors, Eq. (**61**), is presented:

$$\begin{gathered} BP = 968.64(\pm 463.22) + 41.72(\pm 1.16)[59] - 487.65(\pm 226.06)[34] - \\ - 187.94(\pm 30.38)[22] - 2.84(\pm 1.45)[47] \\ n = 45 \qquad r = 0.997 \qquad s = 3.4\ {}^{\circ}C \qquad F = 1696.7 \end{gathered} \tag{61}$$

Correlations for the Acyclic C_1-C_4 Halogenated Compounds (Hal Set)

The correlations for the **Hal** set of compounds were developed with a procedure similar to that used for the **O-S** set. Table 3 presents the notation of the descriptors [64] to [102] involved in the QSPR correlations for the **Hal** set; the descriptors already defined in Table 1 are not included here. A selection of the correlations obtained for the whole set of 534 haloalkanes is presented in Table 4, Eqs. (**76-100**). Due to the higher diversity of the compounds from the **Hal** set, the mono-parametric correlations are of lower quality than those obtained for the **O-S** set, with a large standard deviation, between 32.1 and 45.3 °C. Four topological and one quantum indices are present in the mono-parametric Eqs. (**76-80**): ${}^1\chi^v$, ϕ, the flexibility index, ${}^0\chi^v$, ${}^2\chi^v$, and FNSA-3, the fractional negative surface area computed with quantum charges. The introduction of a second parameter improves the statistical indices for the **Hal** set, and the standard deviation decreases below 17 °C, as can be seen from Eqs. (**81-85**). In Eq. (**81**) the ${}^1\chi^v$ index appears together with the relative number of F atoms, **NoF/N**. A group of three descriptors appears in all equations with 5 and 6 descriptors: [20], the molecular volume, [16], the topographic electronic index, [94], the Kier flexibility index ϕ. The best equation, Eq. (**91**), contains also descriptor [1], the molecular weight, and [81], maximum electron-nuclear attraction for a C atom:

$$\begin{gathered} BP = -266.09(\pm 12.55) + 2.01(\pm 0.03)[20] - 146.27(\pm 1.80)[16] + \\ + 11.50(\pm 0.53)[94] + 0.1617(\pm 0.0083)[1] + 0.9031(\pm 0.0700)[81] \\ n = 534 \qquad r = 0.990 \qquad s = 9.0\ {}^{\circ}C \qquad F = 5281.6 \end{gathered} \tag{91}$$

Table 2. Statistical Indices *r*, *s*, F, and Structural Descriptors Involved in the Multi-Linear Regression Equations Containing Up to Six Parameters Developed to Compute the Boiling Points of Oxygen and Sulfur Containing Compounds

Eq.	Descriptors	Type	*r*	*s*	F
	Compounds with O, S (n = 185)				
1	59	T	0.961	13.4	2211
2	25	Q	0.922	18.6	1049
3	1	C	0.918	19.1	989
4	26	Q	0.887	22.3	678
5	60	T	0.872	23.7	580
6	58 2	TC	0.979	9.9	2086
7	59 9	TE	0.978	10.1	2001
8	59 6	TE	0.978	10.2	1953
9	59 7	TE	0.977	10.2	1950
10	59 8	TE	0.975	10.6	1799
11	59 9 12 13	TEEE	0.987	7.9	1669
12	58 2 22 14	TCGE	0.986	8.1	1600
13	58 2 22 56	TCGQ	0.986	8.3	1521
14	59 9 12 22	TEEG	0.985	8.4	1486
15	59 6 22 57	TEGQ	0.985	8.4	1485
16	59 9 12 13 14	TEEEE	0.989	7.2	1624
17	59 6 54 55 15	TEQQE	0.989	7.3	1557
18	59 6 54 55 16	TEQQE	0.988	7.4	1523
19	59 6 54 13 17	TEQEE	0.988	7.4	1517
20	59 6 54 55 24	TEQQQ	0.988	7.4	1513
21	59 9 12 13 17 39	TEEEEQ	0.992	6.3	1765
22	59 9 12 13 17 3	TEEEEC	0.991	6.4	1705
23	59 9 12 13 17 18	TEEEEE	0.991	6.5	1668
24	59 9 12 13 17 10	TEEEEE	0.991	6.6	1620
25	59 9 12 13 17 2	TEEEEC	0.991	6.6	1587
	Monoethers (n = 73)				
26	58	T	0.988	7.0	2961
27	15	E	0.987	7.2	2753
28	59	T	0.983	8.4	2032
29	20	G	0.980	9.1	1720
30	21	G	0.979	9.2	1678
31	58 27	TQ	0.992	5.7	2207
32	58 28	TQ	0.992	5.7	2192
33	21 29	GQ	0.992	5.8	2175
34	58 30	TQ	0.992	5.9	2051
35	58 31	TQ	0.991	5.9	1923
36	58 27 22 36	TQGQ	0.996	4.0	2227
37	58 27 22 36 37	TQGQ	0.996	4.1	2124
38	58 27 44 39	TQQQ	0.996	4.2	2110
39	58 28 22 36	TQGQ	0.996	4.2	2055
40	58 27 22 44	TQGQ	0.996	4.2	2039
41	58 27 44 39 61	TQQQT	0.997	3.9	1963
42	58 27 22 36 40	TQGQQ	0.997	3.9	1935
43	58 27 44 39 41	TQQQQ	0.997	3.9	1918
44	58 27 22 36 42	TQGQQ	0.997	3.9	1916
45	58 27 22 36 43	TQGQQ	0.997	3.9	1914
46	58 27 39 61 36 19	TQQTQE	0.997	3.8	1697
47	58 27 44 39 61 45	TQQQTQ	0.997	3.8	1692
48	58 27 44 39 61 41	TQQQTQ	0.997	3.8	1689
49	58 27 44 39 61 46	TQQQTQ	0.997	3.8	1679
50	58 27 22 36 42 29	TQGQQQ	0.997	3.8	1678
	Monosulfides (n = 45)				
51	59	T	0.992	5.3	2809
52	58	T	0.991	5.6	2474
53	23	Q	0.989	6.4	1936
54	24	Q	0.987	7.0	1603
55	21	G	0.986	7.2	1477
56	59 32	TQ	0.995	4.2	2208
57	59 33	TQ	0.995	4.4	2029
58	59 34	TQ	0.994	4.7	1809
59	59 5	TE	0.994	4.9	1667
60	59 35	TQ	0.994	4.9	1633
61	59 34 22 47	TQGQ	0.997	3.4	1697
62	59 34 22 10	TQGE	0.997	3.4	1696
63	59 34 22 11	TQGE	0.997	3.4	1670
64	59 34 22 49	TQGQ	0.997	3.5	1669
65	59 34 22 9	TQGE	0.997	3.5	1664
66	59 34 22 47 37	TQGQQ	0.998	3.2	1575
67	59 34 22 47 48	TQGQQ	0.997	3.2	1540
68	59 34 22 47 9	TQGQE	0.997	3.2	1518
69	59 34 22 47 38	TQGQQ	0.997	3.3	1495
70	59 34 22 47 62	TQGQT	0.997	3.3	1479
71	59 34 50 9 63 47	TQQETQ	0.998	3.0	1481
72	59 32 50 35 51 52	TQQQQQ	0.998	3.1	1415
73	59 34 22 47 37 53	TQGQQQ	0.998	3.1	1412
74	59 34 22 47 37 62	TQGQQT	0.998	3.1	1395
75	59 34 22 47 37 63	TQGQQT	0.998	3.1	1392

Table 3. **Notation of the Descriptors Involved in the QSPR Equations from Table 4 that Compute the Boiling Points of Halogen Containing Compounds.** The Following Descriptors That are Used in Table 4 Were Defined in Table 2: [1], [6], [7], [8], [12], [16], [20], [21], [24], [25], [27], [45], [49], [58], [59], [60], [63]

No.	Type	Descriptors
[64]	C	Relative number of F atoms
[65]	C	Number of F atoms
[66]	C	Gravitation index (all bonds)
[67]	C	Relative molecular weight
[68]	C	Number of Cl atoms
[69]	C	Relative number of Cl atoms
[70]	C	Relative number of C atoms
[71]	C	Number of H atoms
[72]	E	DPSA-3 Difference in CPSAs (PPSA3-PNSA3) (electrostatic)
[73]	E	FNSA-1 Fractional PNSA (PNSA-1/TMSA) (electrostatic)
[74]	E	FNSA-3 Fractional PNSA (PNSA-3/TMSA) (electrostatic)
[75]	E	PNSA-3 Atomic charge weighted PNSA (electrostatic)
[76]	E	WPSA-3 Weighted PPSA (PPSA3*TMSA/1000) (electrostatic)
[77]	E	Maximum partial charge for a H atom (electrostatic)
[78]	G	Moment of inertia B
[79]	Q	RNCG Relative negative charge (QMNEG/QTMINUS) (quantum)
[80]	Q	DPSA-2 Difference in CPSAs (PPSA2-PNSA2) (quantum)
[81]	Q	Maximum electron-nucleus attraction for a C atom
[82]	Q	Maximum σ-σ bond order
[83]	Q	Maximum atomic state energy for a C atom
[84]	Q	WPSA-2 Weighted PPSA (PPSA2*TMSA/1000) (quantum)
[85]	Q	Total molecular two-center resonance energy / number of atoms
[86]	Q	Minimum nucleus-nucleus repulsion for a C-C bond
[87]	Q	FPSA-3 Fractional PPSA (PPSA-3/TMSA) (quantum)
[88]	Q	FNSA-3 Fractional PNSA (PNSA-3/TMSA) (quantum)
[89]	Q	PNSA-3 Atomic charge weighted PNSA (quantum)
[90]	Q	Minimum nucleophilic reactivity index for a F atom
[91]	Q	Minimum one-electron reactivity index for a C atom
[92]	Q	Maximum nucleus-nucleus repulsion for a C-C bond
[93]	Q	Average one-electron reactivity index for a C atom
[94]	T	ϕ, Kier flexibility index
[95]	T	$^{2}\kappa$, Kier shape index (order 2)
[96]	T	$^{3}\chi^{v}$, Kier and Hall index (order 3)
[97]	T	$^{1}\kappa$, Kier shape index (order 1)
[98]	T	^{0}IC, information content (order 0)
[99]	T	^{1}IC, Information content (order 1)
[100]	T	^{0}SIC, structural information content (order 0)
[101]	T	^{1}SIC, structural information content (order 1)
[102]	T	^{1}BIC, bonding information content (order 1)

Because CFCs have a particular importance we have separated all CFCs and performed a QSAR analysis for them. As given in Eqs. **(101-105)** from Table 4, the best mono-parametric correlations are obtained with four Kier and Hall connectivity indices and a constitutional descriptor, the number of Cl atoms. The equation with the highest F value is Eq. **(116)**:

$$\begin{aligned} BP = &- 242.32(\pm 12.57) + 60.93(\pm 1.72)[59] - 39.36(\pm 3.15)[16] + \\ &+ 11.56(\pm 0.64)[7] + 126.54(\pm 9.24)[69] + 469.73(\pm 54.07)[70] \\ n = 281 \quad & r = 0.994 \quad s = 7.3\ ^{\circ}C \quad F = 5022.9 \end{aligned} \tag{116}$$

Another set of correlations was developed with for the 42 fluoroalkanes. The results presented in Table 4, Eqs. **(98-102)**, show that the best descriptors in mono-parametric correlations are 5 chemical information indices. Also, all correlations with 4, 5, and 6 parameters contain the ${}^1\chi^v$ index and positive charge surface area index PPSA-3. The best equation with 4 variables, Eq. **(136)**, is presented:

$$\begin{aligned} BP = &- 169.42(\pm 5.83) + 55.97(\pm 2.93)[59] + 16.21(\pm 0.78)[7] + \\ &+ 0.1820(\pm 0.0435)[27] + 624.51(\pm 199.18)[91] \\ n = 42 \quad & r = 0.985 \quad s = 7.3\ ^{\circ}C \quad F = 296.6 \end{aligned} \tag{136}$$

DISCUSSION

The results obtained for the **O-S** set show that the QSPR approach implemented in CODESSA can generate good structure-property models by screening a large set of descriptors. As pointed before, CODESSA offered as the best bi-parametric model the Eq. **(6)** that is identical with Eq. (A) previously reported by us. On the other hand, the greater diversity of descriptors used by CODESSA ensured the discovery of better QSPR models than those already reported. We have to mention that from the 60 multi-parametric correlation equations presented in Table 2 all but one, namely Eq. **(33)**, contain a connectivity topological index, either ${}^1\chi$ or ${}^1\chi^v$. Moreover, the best descriptor in the mono-parametric equations from Table 2 is the ${}^1\chi^v$ index for the **O-S** and the monosulfides sets, and the ${}^1\chi$ index for the monoethers set, respectively. This finding shows that simple to compute topological indices are very effective in correlating the boiling points of organic compounds. The numerical description of the molecular structure with the aid of topological indices represents an efficient and simple way of generating QSPR and QSAR descriptors especially in the case of acyclic, highly flexible molecules, like those investigated in this paper. The geometry of acyclic molecules presents many local minima, making thus difficult to characterize in a simple and unique way the global structure with geometric or quantum structural descriptors. Also, at the boiling point all molecular conformations, sufficiently close in energy to the global minimum, are populated to some extent depending on the difference to the minimum energy. Only techniques of exploring the conformational space and molecular dynamics simulations can offer information on the population of the low energy conformations. The identification of all conformations and the computation of the molecular population of each minima is computationally expensive, and for these reasons molecular dynamics computations are not routinely used in QSPR studies.

The fact that connectivity indices are well represented in the correlations of the **O-S** set adds a substantial argument in the favor of topological indices use in QSPR, QSAR and similarity/diversity screening of organic compounds. However, topology alone is not the only factor determining the boiling point of a compound, and the presence of geometric, quantum, and charge-surface descriptors in the equations from Table 2 greatly improve the structure-property models, adding information that is not present in the topological description of the molecular structure. The best pure topological models for the **O-S** set, given in Eqs. (C-I), are surpassed by the mixed descriptor correlations offered by CODESSA. Important descriptors suggested by the correlations to determine the boiling point are the charge-surface area descriptors, the molecular surface area and volume, and the relative positive charge.

The boiling points of sulfur-containing chains present interesting trends, as can be observed from the data presented in Table 5. If we compare with the alkanes having the same skeleton, *n*-hexane with BP = 69.0 °C, and *n*-heptane with BP = 98.4 °C, the presence of a sulfur atom greatly increases the boiling point, owing to the

Table 4. Statistical Indices *r*, *s*, F, and Structural Descriptors Involved in the Multi-Linear Regression Equations Containing Up to Six Parameters Developed to Compute the Boiling Points of Halogen Containing Compounds

Eq.	Descriptors	Type	*r*	*s*	F
	Halogenated compounds (n = 534)				
76	59	T	0.865	32.1	1586
77	94	T	0.772	40.7	788
78	60	T	0.761	41.5	735
79	63	T	0.723	44.3	583
80	49	Q	0.707	45.3	534
81	59 64	TC	0.973	14.6	4840
82	59 65	TC	0.970	15.6	4220
83	20 16	GE	0.969	15.9	4070
84	59 66	TC	0.968	16.2	3902
85	59 24	TQ	0.966	16.6	3683
86	20 16 94 1	GETC	0.987	10.3	4999
87	20 16 94 79	GETQ	0.986	10.7	4665
88	20 16 1 95	GECT	0.986	10.7	4658
89	20 16 94 67	GETC	0.986	10.9	4483
90	20 64 72 80	GCEQ	0.985	10.9	4427
91	20 16 94 1 81	GETGQ	0.990	9.0	5281
92	20 16 94 67 82	GETCQ	0.990	9.0	5263
93	20 16 94 67 81	GETCQ	0.990	9.0	5255
94	20 16 94 1 83	GETGQ	0.990	9.1	5178
95	20 16 94 1 82	GETGQ	0.990	9.1	5133
96	20 16 94 67 82 45	GETCQQ	0.991	8.6	4810
97	20 16 94 67 81 45	GETCQQ	0.991	8.6	4796
98	20 16 94 67 81 12	GETCQE	0.991	8.6	4789
99	20 16 94 1 81 73	GETGQE	0.991	8.7	4776
100	20 16 94 1 81 84	GETGQQ	0.991	8.7	4761
	F, Cl (n = 281)				
101	59	T	0.894	31.0	1117
102	68	C	0.877	33.2	937
103	63	T	0.840	37.6	672
104	60	T	0.823	39.4	585
105	96	T	0.813	40.3	547
106	59 16	TE	0.985	12.0	4533
107	65 58	TT	0.985	12.1	4488
108	21 16	GE	0.983	12.6	4087
109	59 85	TQ	0.982	13.2	3722
110	59 97	TT	0.981	13.3	3636
111	59 16 7 69	TEEC	0.993	8.2	4933
112	65 58 74 83	TTEQ	0.993	8.5	4644
113	65 58 75 83	TTEQ	0.993	8.5	4564
114	59 16 6 78	TEEG	0.992	8.6	4476
115	65 58 72 69	TTEC	0.992	8.7	4389
116	59 16 7 69 70	TEEEC	0.994	7.3	5022
117	59 16 7 69 86	TEEEQ	0.994	7.5	4779
118	59 16 6 69 70	TEECC	0.994	7.5	4699
119	65 58 72 69 87	TTECQ	0.994	7.5	4690
120	59 16 7 69 27	TEEEQ	0.994	7.6	4665
121	59 16 6 69 70 88	TEECCQ	0.995	6.9	4610
122	65 58 72 69 87 25	TTECQQ	0.995	7.0	4571
123	59 16 6 69 70 89	TEECCQ	0.995	7.0	4553
124	59 16 6 69 70 90	TEECCQ	0.995	7.0	4545
125	59 16 6 78 69 70	TEEGCC	0.995	7.0	4498
	F (n = 42)				
126	98	T	0.893	18.1	158
127	99	T	0.862	20.3	116
128	100	T	0.843	21.6	98
129	101	T	0.842	21.7	97
130	102	T	0.827	22.6	86
131	59 7	TE	0.976	8.9	390
132	71 75	CE	0.970	9.9	311
133	59 76	TE	0.968	10.2	291
134	21 7	GE	0.968	10.3	287
135	25 7	QE	0.966	10.6	269
136	59 7 27 91	TEQQ	0.985	7.3	296
137	59 7 92 91	TEQQ	0.985	7.3	294
138	59 7 86 91	TEQQ	0.985	7.3	292
139	59 7 27 77	TEQE	0.984	7.5	275
140	59 7 86 77	TEQE	0.983	7.6	270
141	59 7 27 91 77	TEQQE	0.987	6.7	280
142	59 7 86 77 90	TEQEQ	0.987	6.8	275
143	59 7 27 77 90	TEQEQ	0.987	6.8	272
144	59 7 92 77 90	TEQEQ	0.987	6.8	272
145	59 7 27 91 93	TEQQQ	0.987	6.9	267
146	59 7 27 91 77 93	TEQQEQ	0.989	6.3	267
147	59 7 92 77 75 71	TEQEEC	0.989	6.3	264
148	59 7 92 77 72 71	TEQEEC	0.989	6.3	264
149	59 7 86 77 75 71	TEQEEC	0.989	6.4	260
150	59 7 86 77 72 71	TEQEEC	0.989	6.5	260

increased molecular weight.[16] For isomeric sulfides with the same skeleton having a more centrally situated sulfur atom, the boiling point is slightly lower than for compounds with a more marginal one. For isomeric disulfides having the same skeleton the boiling point is higher when the two heteroatoms are farther apart. The above trends are well reproduced by the boiling points computed with Eq. (**21**), as can be observed from the computed values and residuals reported in Table 5.

Table 5. Comparison of Boiling Points (Experimental and Calculated with Eq. (**21**)) for Sulfides and Disulfides Having the Same Skeleton

Name	BP_{exp}	BP_{calc}	BP_{calc}-BP_{exp}
butyl methyl sulfide	123.2	121.6	−1.6
ethyl propyl sulfide	118.5	115.9	−2.6
1,2-bis(methylthio)ethane	183.0	176.4	−6.6
ethylthio-methylthio-methane	166.0	160.2	−5.8
diethyl disulfide	154.0	151.5	−2.5
methyl pentyl sulfide	145.0	145.7	0.7
butyl ethyl sulfide	144.2	139.1	−5.1
bis(ethylthio)methane	181.0	177.7	−3.3
ethyl propyl disulfide	173.7	173.4	−0.3
dipropyl sulfide	142.8	139.1	−3.7

The equations offered by CODESSA for the **Hal** set show a significant improvement over the previous reported correlations, Eqs. (J-L). The best descriptor in the mono-parametric equations from Table 4 is the $^1\chi^v$ index for the **Hal** and the chlorofluorocarbons sets, and the information content IC_0 index for the fluorocarbons, respectively. From the 15 best mono-parametric equations reported in Table 4 for the three sets of halogenated compounds, 13 contain a topological index. We have to mention that in the case of the fluorocarbons all 5 indices from the mono-parametric equations are information indices, showing that for fluorine the usual connectivity indices are not particularly fit in mono-parametric equations. An inspection of the Eqs. (**136-150**) representing correlations with 4, 5, and 6 descriptors for fluorocarbons shows that the $^1\chi^v$ index appears in all 15 equations, together with PPSA-3, descriptor [7], the atomic charge weighted partial positive surface area. The index $^1\chi^v$ appears in 10 equations while $^1\chi$ appears in the remaining 5 equations with 4, 5, and 6 descriptors for chlorofluorocarbons. As already mentioned, for the **Hal** set, in Eqs. (**86-100**) there appears the Kier flexibility index ϕ and no conectivity index.

The correlations containing only constitutional descriptors and topological indices for the **Hal** set, given in Eqs. (J-L), are of slightly lower statistical quality than the mixed descriptor correlations offered by CODESSA. The descriptors that appear in the correlations for the **Hal** set are to a large extent different from those appearing in the **O-S** set, showing that the separation in sets according to the chemical constitution is supported by the different factors influencing the boiling points in the two sets. Also there is a significant improvement in the correlations developed for subsets of constitutionally related compounds, as can be observed by comparing the statistical indices of Eqs. (**21**), (**36**), and (**61**), with the standard deviation s = 6.3 °C for the **O-S** set, s = 4.0 °C for the monoethers set and s = 3.4 °C for the monosulfides set. This improvement is observed for the **Hal** set, as can be seen by comparing Eqs. (**91**), (**116**) and (**136**): the standard deviation decreases from 9.0 °C to 7.3 °C for the chlorofluorocarbons and fluorocarbons sets. The results obtained in this study indicate that for the estimation of the normal boiling points models derived for heterogeneous sets of organic compounds are outperformed by more specific models, each model being developed for a narrow class of compounds.

Some interesting trends can be observed by comparing the boiling points of chloro-, dichloro- and trichlorobutanes presented in Table 6. By comparing isomeric systems, one observes the same trend as for alkanes with the same molecular graph, whose vertices correspond to both carbon and chlorine atoms. Namley, branching lowers boiling points. The above trends are reproduced by the computed BP with Eq. (**91**), as can be seen in Table 6. Comparing the residuals from Table 6 with the standard deviation of Eq. (**91**), s = 9.0 °C, one can see that only the BP of 1,1-dichlorobutane is affected by a large error.

Table 6. Comparison of Boiling Points (Experimental and Calculated with Eq. (**91**)) for Chloro-, Dichloro-, and Trichlorobutanes

Name	BP_{exp}	BP_{calc}	BP_{calc}-BP_{exp}
1,4-dichlorobutane	155.0	156.7	1.7
1,3-dichlorobutane	133.0	132.7	–0.3
1,2-dichlorobutane	123.5	132.3	8.8
2,3-dichlorobutane	116.0	115.9	–0.1
1,1-dichlorobutane	115.0	125.9	10.9
2,2-dichlorobutane	104.0	99.0	–5.0
1,1,4-trichlorobutane	183.8	181.5	–2.3
1,2,3-trichlorobutane	166.0	170.8	4.8
1,1,2-trichlorobutane	156.8	160.2	3.4
1,1,3-trichlorobutane	153.8	161.1	7.3
2,2,3-trichlorobutane	143.0	139.4	–3.6
1,1,1-trichlorobutane	133.5	139.1	5.6

Eq. (**116**) predicts for Cl_2CHCH_2F BP = 70 °C and for Cl_3CCH_2F BP = 85 °C in agreement with literature data.[22a, b]

Acknowledgements

One of the authors (OI) acknowledges a software grant consisting of HyperChem Release 4.5 offered by Hypercube, Inc., and a software grant consisting of AMPAC 5.0 and CODESSA 2.13 offered by Semichem.

REFERENCES

1. Rechsteiner, C. E. In: *Handbook of Chemical Property Estimation Methods*; Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H., Eds.; McGraw-Hill: New York, 1982; Chapter 12.
2. Reid, R. C.; Prausnitz, J. M.; Poling, B. E. *The Properties of Gases and Liquids*, 4th ed.; McGraw-Hill: New York, 1987.
3. Horvath, A. L. *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds*; Elsevier: Amsterdam, 1992.
4. Stuper, A. J.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 99-105; Brugger, W. E.; Stuper, A. J.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 105-110.
5. Hansen, P. J.; Jurs, P. C. *Anal. Chem.* **1987**, *59*, 2322-2327; Stanton, D. T.; Jurs, P. C.; Hicks, M. G. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301-310; Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306-316; Egolf, L. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 616-625; Egolf, L. M.; Wessel, M. D.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 947-956; Wessel, M. D.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 68-76; Wessel, M. D.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 841-850.
6. Mekenyan, O.; Bonchev, D. *Acta Pharm. Jugosl.* **1986**, *36*, 225-237; Mekenyan, O.; Karabunarliev, S.; Bonchev, D. *Computers Chem.* **1990**, *14*, 193-200; Mekenyan, O. G.; Karabunarliev, S. H.; Ivanov, J. M.; Dimitrov, D. N. *Comput. Chem.* **1994**, *18*, 173-187.
7. Bonchev, D.; Mountain, C. F.; Seitz, W. A.; Balaban, A. T. *J. Med. Chem.* **1993**, *36*, 1562-1569; Bonchev, D.; Seitz, W. A.; Mountain, C. F.; Balaban, A. T. *J. Med. Chem.* **1994**, *37*, 2300-2307.
8. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, 279-287; Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027-1043.
9. Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. *Anal. Chem.* **1994**, *66*, 1799-1807; Murugan, R.; Grendze, M. P.; Toomey, Jr., J. E.; Katritzky, A. R.; Karelson, M.; Lobanov, V. S.; Rachwal, P. *CHEMTECH* **1994**, *24*, 17-23; Katritzky, A. R.; Rachwal, P.; Law, K. W.; Karelson, M.; Lobanov, V. S. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 879-884; Katritzky, A. R.; Mu, L.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1162-1168; Katritzky, A. R.; Lobanov, V. S.;

Karelson, M.; Murugan, R.; Grendze, M. P.; Toomey, Jr. J. E. *Rev. Roum. Chim.* **1996**, *41*, 851-867; Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. *Langmuir* **1996**, *12*, 1462-1470; Katritzky, A. R.; Mu, L.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 756-761; Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 913-919; Katritzky, A. R.; Mu, L.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 293-299.

10. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem.* **1996**, *100*, 10400-10407.
11. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28-41.
12. Cramer, R. D. III. *J. Am. Chem. Soc.* **1980**, *102*, 1837-1849.
13. Abramowitz, R.; Yalkowsky, S. H. *Pharm. Res.* **1990**, *7*, 942-947; Devotta, S.; Pendyala, V. R. *Ind. Eng. Chem. Res.* **1992**, *31*, 2042-2046; Simamora, P.; Miller, A. H.; Yalkowsky, S. H. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 437-440; Simamora, P.; Yalkowsky, S. H. *SAR QSAR Environ. Res.* **1993**, *1*, 293-300; Simamora, P.; Yalkowsky, S. H. *Ind. Eng. Chem. Res.* **1994**, *33*, 1405-1409; Yalkowsky, S. H.; Krzyzaniak, J. F.; Myrdal, P. B. *Ind. Eng. Chem. Res.* **1994**, *33*, 1872-1877; Stein, S. E.; Brown, R. L. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 581-587; Krzyzaniak, J. F.; Myrdal, P. B.; Tsibanogiannis, I. N.; Kalospiros, N. S.; Tassios, D. P. *Ind. Eng. Chem. Res.* **1995**, *34*, 997-1002; Simamora, P.; Yalkowsky, S. H. *Ind. Eng. Chem. Res.* **1995**, *34*, 2530-2535.
14. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
15. White, C. M. *J. Chem. Eng. Data* **1986**, *31*, 198-203; Needham, D. E.; Wei, I. C.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 4186-4194; Gautzsch, R.; Zinn, P. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 791-800; Wang, S.; Milne, G. W. A.; Klopman, G. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1242-1250; Basak, S. C.; Gute, B. D.; Grunwald, G. D. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1054-1060; Estrada, E. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23-27; Nikolić, S.; Trinajstić, N.; Baučić, I. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 42-46.
16. Balaban, A. T.; Kier, L. B.; Joshi, N. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 237-244.
17. Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 000.
18. Balaban, A. T.; Joshi, N.; Kier, L. B.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 233-237.
19. Basak, S. C.; Gute, B. D.; Grunwald, G. D. *Croat. Chem. Acta* **1996**, *69*, 1159-1173.
20. Le, T. D.; Weers, J. G. *J. Phys. Chem.* **1995**, *99*, 13909-13916; Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, M. E.; Politzer, P. *J. Phys. Chem.* **1993**, *97*, 9369-9373.
21. Cherqaoui, D.; Villemin, D. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 97-102; Hall, L. H.; Story, C. T. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004-1014.
22. a. Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. D. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118-1121; b. Carlton, T.S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 158-164.
23. Cherqaoui, D.; Villemin, D.; Mesbah, A.; Cense, J. M.; Kvasnicka, V. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 2015-2019.
24. Lohninger, H. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 736-744.
25. Ivanciuc, O. MolNet Neural Network Application in Structure-Property Studies, The 23rd Chemistry Conference, 8-10 October **1997**, Călimăneşti, Vâlcea, Romania.
26. CRC Handbook of Physics and Chemistry, 68th ed.; CRC: Boca Raton, FL.
27. Dictionary of Organic Compounds, 5th ed.; Chapman and Hall: New York, **1982**.
28. Smith, B. D.; Srivastava, R. *Thermodynamic Data for Pure Compounds, Part B; Halogenated Hydrocarbons and Alcohols, Physical Sciences Data No.* 25; Elsevier: Amsterdam, 1986.
29. HyperChem 4.5, Hypercube, Inc., 419 Phillip St., Waterloo, Ontario, Canada N2L 3X2; Tel. (519)-725-4040; Fax: (519)-725-5193; information hot-line: (800)-960-1871; E-mail information requests: info@hyper.com; URL: http://www.hyper.com.
30. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
31. AMPAC 5.0, Semichem, 7204 Mullen, Shawnee, KS 66216, USA, E-mail: andy@semichem.com, http://www.semichem.com.
32. CODESSA 2.13, Semichem, 7204 Mullen, Shawnee, KS 66216, USA, E-mail: andy@semichem.com, http://www.semichem.com.