

# Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure–Property Relationship

Alan R. Katritzky,<sup>\*,†</sup> Victor S. Lobanov,<sup>‡</sup> and Mati Karelson<sup>§</sup>

Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P.O. Box 117200, Gainesville, Florida 32611-7200, 3-Dimensional Pharmaceuticals, Inc., 665 Stockton Drive, Suite 104, Exton, Pennsylvania 19341, and Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu EE 2400, Estonia

Received May 15, 1997<sup>®</sup>

We recently reported a successful correlation of the normal boiling points of 298 organic compounds containing O, N, Cl, and Br with two molecular descriptors.<sup>1</sup> In the present study the applicability of these two descriptors for the prediction of boiling points for various other classes of organic compounds was investigated further by employing a diverse data set of 612 organic compounds containing C, H, N, O, S, F, Cl, Br, and I. The data set was divided into 9 subsets, and additional descriptors were sought for each subset, which, together with the gravitation index and the charged surface area of hydrogen-donor atoms, would model the boiling points. The additional descriptors were then each tested for global relevance and retained only if this was found. A final eight-parameter correlation model was thus deduced which had  $R^2 = 0.965$  and a standard error of 15.5 K approaching the estimated average experimental error for the data set. The model appears to be general for a wide variety of organic compounds.

## INTRODUCTION

The normal boiling point (i.e. the boiling point at 1 atm) is one of the major physicochemical properties used to characterize and identify a compound. Besides being an indicator for the physical state (liquid or gas) of a compound, the boiling point also provides an indication of its volatility. The degree of volatility is particularly important in industrial applications, as the hazardous nature of a chemical is usually closely related to its volatility. Moreover, boiling points can be used to predict or estimate other physical properties, such as critical temperatures,<sup>2</sup> flash points,<sup>3</sup> enthalpies of vaporization,<sup>4</sup> etc. The boiling point is often the first property measured for a new compound and one of the few parameters known for almost every volatile compound. Normal boiling points are easy to determine; however, when a chemical is unavailable, as yet unknown, or hazardous to handle, a reliable procedure for estimating its boiling point is required. Furthermore, the rapid growth of combinatorial chemistry, where literally millions of new compounds are synthesized and tested without isolation, could render such a procedure very useful.

A large number of methods for estimating boiling points have been devised and numerous quantitative structure–property relationship (QSPR) correlations of normal boiling points have been reported. Earlier methods for boiling point estimation have been summarized by Rechsteiner<sup>4</sup> and Horvath.<sup>5</sup> A detailed review of previous correlations of boiling points has been given by us elsewhere,<sup>1</sup> and additional references on methods for estimating boiling points are also

available in recent reports on boiling point correlations by Wessel and Jurs<sup>6</sup> and Le and Weers.<sup>7</sup>

Table 1 summarizes the best QSPR models previously reported for predicting boiling points. These models provide satisfactory predictions for various classes of organic compounds, but, as discussed below, a more general model is desirable. The general group contribution method by Stein and Brown has the large standard error of 24.6K.<sup>8a</sup> Wang *et al.*<sup>8b</sup> significantly improve this by combining the group contribution approach with molecular weight and local graph indices to obtain, for a diverse set of 541 organic compounds,  $R^2 = 0.992$  and  $SD = 7.57$  K; however, 99 parameters were used. A comparative analysis of the correlation equations of Table 1 reveals that the QSPR models for different classes of compounds have few overlapping descriptors. In addition, many of those models employ simple counts of atoms or groups of specific types and indicator variables which are data-set dependent. The failure to devise a general QSPR model for the prediction of the boiling points of organic compounds is due to the inability of the available descriptors to reflect quantitatively variations in the intermolecular interactions in liquid media.

Recently, we have reported a two descriptor QSPR model that successfully correlates ( $R^2 = 0.954$ ) the normal boiling points of 298 organic compounds containing O, N, Cl, and Br. The two descriptors involved in this model were (i) the cubic root of the gravitation index ( $G_1^{1/3}$ ) and (ii) the charged surface area of the hydrogen-donor atoms (HDSA2).<sup>1</sup> The first descriptor accounts successfully for most of the variation in the boiling points of hydrocarbons and halogenated hydrocarbons, whereas the second descriptor effectively quantifies hydrogen bonding effects. Both descriptors have a clear physical meaning. The cubic root of the gravitation index is a size-dependent descriptor of a compound which

<sup>†</sup> Florida Center for Heterocyclic Compounds.

<sup>‡</sup> 3-Dimensional Pharmaceuticals, Inc.

<sup>§</sup> University of Tartu.

<sup>®</sup> Abstract published in *Advance ACS Abstracts*, November 1, 1997.

**Table 1.** Reported QSPR Models for Predicting Normal Boiling Points of Organic Compounds

compounds	$R^2$	$s$	$n^a$	$m^b$	ref
alkanes	0.999	1.9	74	RA, 5	Needham <i>et al.</i> <sup>27</sup>
hydrocarbons	0.994	6.3	296	RA, 6	Wessel and Jurs <sup>28</sup>
furans, tetrahydrofurans (THFs)	0.968	11.2	209	RA, 11	Stanton <i>et al.</i> <sup>29</sup>
thiophenes	0.974	7.9	134	RA, 7	Stanton <i>et al.</i> <sup>29</sup>
pyrans	0.978	10.2	146	RA, 7	Stanton <i>et al.</i> <sup>21</sup>
pyrroles	0.962	12.3	278	RA, 7	Stanton <i>et al.</i> <sup>21</sup>
furans, THFs, thiophenes, pyrans, and pyrroles	0.954	13.1	752	RA, 11	Stanton <i>et al.</i> <sup>21</sup>
pyridines	0.933	15.0	291	RA, 7	Egolf and Jurs <sup>30</sup>
haloalkanes C <sub>1</sub> –C <sub>4</sub>	0.970	10.9	532	RA, 6	Balaban <i>et al.</i> <sup>23</sup>
haloalkanes C <sub>1</sub> –C <sub>4</sub>	0.992	8.5	276	NN	Balaban <i>et al.</i> <sup>31</sup>
fluorocarbons	0.985	9.5	68	RA, 3	Le and Weers <sup>7</sup>
acyclic ethers, peroxides, acetals, and their sulfur analogues	0.971	8.2	185	RA, 3	Balaban <i>et al.</i> <sup>32</sup>
rigid aromatic compounds	N/A	17.3	246	GC, 22	Simamora <i>et al.</i> <sup>33</sup>
O, N, Cl, and Br containing compounds	0.976	11.8	268	RA, 8	Egolf <i>et al.</i> <sup>26</sup>
O, S, and halogens containing compounds	0.982	11.6	248	RA, 10	Wessel and Jurs <sup>6</sup>
N containing compounds	0.980	10.7	90	RA, 10	Wessel and Jurs <sup>6</sup>
O, N, Cl, and Br containing compounds	0.973	12.4	298	RA, 4	Katritzky <i>et al.</i> <sup>1</sup>
diverse organic compounds	N/A	24.6	4426	GC, 85	Stein and Brown <sup>8a</sup>
diverse organic compounds	0.965	15.5	584	RA, 8	present paper

<sup>a</sup> Number of compounds used to devise a model. <sup>b</sup> Technique utilized (RA, regression analysis; GC, group contributions approach; NN, neural networks) and number of parameters in the model.

accounts for the dispersion and cavitation effects in the liquid media.<sup>9</sup> The gravitation index itself, as discussed in our previous paper,<sup>1</sup> is calculated as follows:

$$G_1 = \sum_{i < j}^{N_B} \frac{m_i m_j}{r_{ij}^2} \quad (1)$$

where  $m_i$  denotes an atomic mass of the  $i$ th atom,  $r_{ij}$  denotes a distance between  $i$ th and  $j$ th atoms, and summation is done over all bonded pairs of atoms in the molecule. The HDSA2 descriptor is directly related to the hydrogen bonding between the molecules in condensed media. As also described in our previous paper,<sup>1</sup> it is calculated according to the following formula:

$$\text{HDSA2} = \sum_D \frac{q_D S_D^{1/2}}{S_{\text{tot}}} \quad (2)$$

where  $q_D$  is the partial charge on the hydrogen bonding donor atom,  $S_D$  is the surface area of this atom, and  $S_{\text{tot}}$  is the total surface area of the molecule. The summation in the last formula is performed over all possible hydrogen bonding donor sites in the molecule.

An extended four-parameter model ( $R^2 = 0.973$ ) provided a better standard error of 12.4 K; however, the boiling points of some nitrogen containing compounds were still predicted less satisfactorily. For example, for the nitriles, the difference between the experimental and predicted values of normal boiling points was on average larger than 20 K. A similar problem was reported by Wessel and Jurs who developed two separate QSPR models for calculating boiling points for non-nitrogen and for nitrogen containing compounds.<sup>6</sup> In the present study we examine the applicability of the two descriptors mentioned above for the prediction of boiling points for various other classes of organic compounds and investigate the additional factors needed to devise a more general QSPR model for the normal boiling points of organic compounds.

## RESULTS AND DISCUSSION

**Data Set.** The data set of 298 organic compounds containing O, N, Cl, and Br from our previous study<sup>1</sup> was combined with the two sets of 277 compounds containing only O, S, and halogens and 104 compounds all containing N from the paper by Wessel and Jurs.<sup>6</sup> These two data sets had a significant number of overlapping compounds with our original data set, so the resulting compilation included 525 compounds. Finally, 87 boiling points were selected from the *CRC Handbook of Chemistry and Physics*<sup>10</sup> and the *Aldrich Catalog Handbook of Fine Chemicals*<sup>11</sup> to increase the presence and diversity of O, N, S, F, Cl, Br, and I containing compounds in the set (namely, ketones, nitro compounds, nitriles, sulfones, sulfites, perfluorinated compounds, polyhalogenated compounds, and iodinated compounds). These additional compounds were selected so that each of the 9 subsets described below would have at least 30 boiling points. The final set of 612 compounds is representative for all major classes of organic compounds containing C, H, O, N, S, F, Cl, Br, and I (cf. Table 2).

**Descriptor Calculation and Regression Analysis.** All structures were drawn and optimized using the molecular mechanics MMX method within the framework of the PCMODEL<sup>12</sup> program. The preoptimized structures were submitted to AMPAC 5.0 program<sup>13</sup> for further geometry refinement and for the calculation of molecular geometry and wave function related descriptors using the semiempirical AM1<sup>14</sup> parametrization. The output files from AMPAC were transferred to the CODESSA software<sup>15</sup> to calculate various descriptors. The procedures for the calculation of a large selection of molecular descriptors including a variety of constitutional, topological, geometric, and electrostatic descriptors are implemented in CODESSA.<sup>16</sup> The quantum-chemical descriptors extracted from the output of the molecular orbital (AMPAC<sup>13</sup> or MOPAC<sup>17</sup>) calculations include the following: Mulliken net atomic charges, the total dipole moment of the molecule and its components, the frontier molecular orbital (FMO) energies and the respective FMO reactivity indices, molecular polarizability terms, bond orders, energy partitioning terms, etc.<sup>16,18,19</sup> The quantum-

**Table 2.** Experimental and Predicted Boiling Points for 584 Organic Compounds Used in the Study

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
1,1,2,2-tetrabromoethane	516.7	514.4	-2.3	BrI	<i>a</i>
1,1-dibromoethane	381.1	390.3	9.2	BrI	<i>a</i>
1,2-dibromobenzene	497.1	501.2	4.0	BrI	Aldrich
1,2-dibromoethane	404.5	398.5	-6.0	BrI	<i>a</i>
1,3,5-tribromobenzene	544.1	550.3	6.2	BrI	Aldrich
1,3-dibromobenzene	491.6	500.6	9.0	BrI	Aldrich
1,4-diiodobenzene	558.1	550.6	-7.5	BrI	Aldrich
1-bromobutane	374.8	366.7	-8.1	BrI	<i>b</i>
1-bromopropane	344.1	340.0	-4.1	BrI	<i>b</i>
1-iodobutane	403.1	396.3	-6.8	BrI	Aldrich
1-iodoheptane	477.1	464.2	-12.9	BrI	Aldrich
1-iodohexane	452.1	442.8	-9.3	BrI	Aldrich
1-iodopentane	427.1	420.4	-6.8	BrI	Aldrich
2-bromobenzaldehyde	503.1	514.9	11.7	BrI	Aldrich
2-bromobutane	364.4	363.0	-1.4	BrI	<i>b</i>
2-bromoethyl acetate	432.1	421.6	-10.6	BrI	Aldrich
2-bromopropane	332.6	336.1	3.5	BrI	<i>b</i>
2-iodobutane	392.1	393.9	1.8	BrI	Aldrich
2-iodotoluene	484.1	490.0	5.9	BrI	Aldrich
3-bromoanisole	496.1	478.7	-17.4	BrI	Aldrich
3-bromofuran	376.1	418.8	42.6	BrI	Aldrich
4-bromotoluene	457.1	462.8	5.7	BrI	Aldrich
4-iodotoluene	484.6	490.2	5.5	BrI	Aldrich
benzyl bromide	471.6	466.9	-4.8	BrI	Aldrich
bromobenzene	429.2	442.7	13.4	BrI	<i>b</i>
bromoethane	311.5	309.4	-2.1	BrI	<i>a</i>
bromomethyl methyl ether	360.1	340.2	-20.0	BrI	Aldrich
dibromomethane	370.1	370.6	0.5	BrI	<i>a</i>
ethyl iodide	345.4	344.4	-1.0	BrI	<i>a</i>
iodobenzene	461.1	471.4	10.3	BrI	Aldrich
isopropyl iodide	362.6	368.7	6.1	BrI	<i>a</i>
methyl bromide	276.7	278.6	1.9	BrI	<i>a</i>
methyl iodide	315.6	316.8	1.2	BrI	<i>a</i>
<i>n</i> -propyl iodide	375.6	371.7	-3.9	BrI	<i>a</i>
1,1,1-trifluoroacetone	295.1	320.5	25.3	F	Aldrich
1,1,1-trifluoroethane	225.8	227.9	2.1	F	<i>a</i>
1,1,2,2-tetrafluoroethane	250.1	227.2	-22.9	F	<i>a</i>
1,1-difluoroethane	247.4	227.6	-19.8	F	<i>a</i>
1,1-difluoroethylene	187.5	211.5	24.0	F	<i>a</i>
1,2-difluorobenzene	365.1	374.9	9.8	F	Aldrich
1,2-difluoroethane	283.6	237.4	-46.2	F	<i>a</i>
1,3-difluorobenzene	355.1	375.2	20.0	F	Aldrich
1-fluorohexane	365.6	356.7	-9.0	F	Aldrich
2,2,2-trifluoroethyl trifluoroacetate	328.1	335.1	6.9	F	Aldrich
2,2,3,3-tetrafluorocyclobutanitrile	421.1	449.4	28.2	F	Aldrich
2,3,4,5-tetrafluorobenzonitrile	437.6	485.8	48.2	F	Aldrich
3-fluoro- <i>o</i> -xylene	423.1	421.4	-1.7	F	Aldrich
3-fluoropyridine	380.6	395.8	15.2	F	Aldrich
4-fluorotoluene	389.1	400.1	10.9	F	Aldrich
α,α,α-trifluorotoluene	375.1	399.2	24.1	F	Aldrich
carbon tetrafluoride	145.1	126.4	-18.7	F	<i>a</i>
difluoromethane	221.5	160.1	-61.4	F	<i>a</i>
ethyl fluoride	235.4	217.2	-18.2	F	<i>a</i>
fluorobenzene	358.1	374.9	16.7	F	Aldrich
hexafluorobenzene	353.4	352.5	-0.9	F	<i>a</i>
hexafluoroethane	194.9	197.6	2.7	F	<i>a</i>
methyl fluoride	194.8	157.3	-37.5	F	<i>a</i>
2',3',4',5',6'-pentafluoroacetophenone	403.6	464.4	60.8	F	Aldrich
2',3',4',5',6'-pentafluoroanisole	411.6	411.9	0.3	F	Aldrich
pentafluorobenzene	358.1	359.6	1.5	F	Aldrich
pentafluoroethane	225.1	212.2	-12.9	F	<i>a</i>
2,2,2-trifluoroethyl ether	335.6	335.3	-0.4	F	Aldrich
(trifluoromethoxy)benzene	375.1	414.9	39.7	F	Aldrich
vinyl fluoride	200.9	214.8	13.9	F	<i>a</i>
1,1,1,2-tetrachloroethane	403.7	402.9	-0.8	Cl	<i>a</i>
1,1,1-trichloroacetone	445.1	427.9	-17.3	Cl	Aldrich
1,1,1-trichloroethane	347.2	364.8	17.6	Cl	<i>a</i>
1,1,2,2-tetrachloroethane	418.3	412.2	-6.1	Cl	<i>a</i>
1,1,2-trichloroethane	387.0	390.3	3.3	Cl	<i>a</i>
1,1-dichloroethane	330.4	338.0	7.6	Cl	<i>a</i>
1,1-dichloropropane	361.3	366.0	4.7	Cl	<i>b</i>
1,2,3-trichloropropane	430.0	433.9	3.9	Cl	<i>a</i>
1,2,4-trichlorobenzene	486.2	467.9	-18.3	Cl	<i>a</i>
1,2-dichlorobutane	397.1	408.2	11.1	Cl	<i>a</i>
1,2-dichloroethane	356.6	351.0	-5.6	Cl	<i>a</i>
1,2-dichloropropane	369.5	376.9	7.4	Cl	<i>a</i>
1,2-dichloropropene	361.3	350.5	-10.8	Cl	<i>b</i>
1-chloropentane	381.5	375.8	-5.7	Cl	<i>b</i>
2,2,2-trichloroethyl chloroformate	444.6	443.4	-1.2	Cl	Aldrich

Table 2 (Continued)

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
2,3-dichlorobutane	392.6	399.2	6.6	Cl	<i>a</i>
1,3-dichloropropane	393.6	381.1	-12.4	Cl	<i>b</i>
1,4-dichlorobutane	427.1	416.2	-10.9	Cl	<i>b</i>
1,5-dichloropentane	453.1	439.3	-13.8	Cl	<i>b</i>
2-chloroethyl isocyanate	408.1	407.2	-0.9	Cl	Aldrich
2-chloromethyl-1,3-dioxolane	430.6	419.3	-11.3	Cl	Aldrich
2-chloropropene	295.8	304.2	8.4	Cl	<i>a</i>
2-chloroacrylonitrile	361.6	410.5	48.9	Cl	Aldrich
3-chloropropene	318.1	321.9	3.7	Cl	<i>b</i>
4-chlorobenzoyl chloride	495.1	516.5	21.3	Cl	Aldrich
benzotrichloride	486.7	497.5	10.8	Cl	<i>a</i>
benzyl chloride	452.6	454.8	2.2	Cl	<i>a</i>
benzyl dichloride	487.0	486.7	-0.3	Cl	<i>a</i>
carbon tetrachloride	349.8	345.0	-4.8	Cl	<i>a</i>
chloroacetaldehyde diethyl acetal	430.1	447.4	17.3	Cl	Aldrich
chloroacetaldehyde dimethyl acetal	402.1	404.9	2.8	Cl	Aldrich
chloroacetone	393.1	373.7	-19.5	Cl	Aldrich
chloroacetoneitrile	398.1	392.6	-5.6	Cl	Aldrich
chlorobenzene	404.9	408.9	4.0	Cl	<i>b</i>
chloroform	334.3	334.4	0.1	Cl	<i>a</i>
dichloromethane	312.9	305.6	-7.3	Cl	<i>a</i>
ethyl chloride	285.4	280.6	-4.8	Cl	<i>a</i>
ethyl chloroacetate	416.1	392.9	-23.2	Cl	Aldrich
ethyl chloroformate	366.1	349.1	-17.0	Cl	Aldrich
hexachlorobenzene	582.6	576.2	-6.4	Cl	<i>a</i>
isobutyl chloride	342.0	344.6	2.6	Cl	<i>a</i>
isopropyl chloride	308.9	314.8	5.9	Cl	<i>b</i>
<i>m</i> -dichlorobenzene	446.2	437.7	-8.5	Cl	<i>a</i>
methyl chloride	248.9	238.8	-10.1	Cl	<i>a</i>
<i>n</i> -butyl chloride	351.6	348.0	-3.6	Cl	<i>a</i>
<i>n</i> -propyl chloride	319.7	316.4	-3.3	Cl	<i>b</i>
<i>o</i> -dichlorobenzene	453.6	437.4	-16.2	Cl	<i>b</i>
<i>p</i> -dichlorobenzene	447.2	439.9	-7.3	Cl	<i>a</i>
pentachloroethane	433.0	419.5	-13.5	Cl	<i>a</i>
propionyl chloride	351.6	372.8	21.2	Cl	Aldrich
<i>sec</i> -butyl chloride	341.3	345.8	4.5	Cl	<i>b</i>
<i>tert</i> -butyl chloride	323.8	342.2	18.4	Cl	<i>b</i>
trichloroacetoneitrile	356.6	355.0	-1.7	Cl	Aldrich
trichloroacetyl chloride	388.1	422.7	34.6	Cl	Aldrich
1-(2-aminoethyl)piperidine	459.2	448.5	-10.7	NH	<i>a</i>
1-(2-aminoethyl)piperizine	493.2	494.7	1.5	NH	<i>a</i>
1-ethylpiperidine	404.2	402.8	-1.4	NH	<i>a</i>
1-naphthylamine	573.8	555.2	-18.6	NH	<i>a</i>
2-aminoheptane	416.2	410.7	-5.5	NH	<i>a</i>
2-butylamine	335.9	345.0	9.1	NH	<i>a</i>
2-ethylpiperidine	416.2	433.4	17.2	NH	<i>a</i>
2-methyl aniline	473.5	465.7	-7.8	NH	<i>a</i>
2-methylbutylamine	368.7	374.3	5.6	NH	<i>a</i>
2-methylpyrrole	420.7	444.8	24.1	NH	<i>a</i>
2-naphthylamine	579.3	560.0	-19.3	NH	<i>a</i>
2-propylamine	304.9	319.6	14.7	NH	<i>a</i>
3,3-dimethylpiperidine	410.2	440.7	30.5	NH	<i>a</i>
3-methyl aniline	476.5	480.1	3.6	NH	<i>a</i>
allylamine	326.4	337.6	11.2	NH	<i>b</i>
aniline	457.6	449.9	-7.7	NH	<i>b</i>
benzylamine	457.7	457.2	-0.5	NH	<i>a</i>
cyclohexylamine	407.6	413.9	6.3	NH	<i>b</i>
cyclopentylamine	380.2	395.3	15.1	NH	<i>a</i>
di- <i>n</i> -propylamine	382.0	393.2	11.2	NH	<i>b</i>
diamylamine	476.1	474.0	-2.1	NH	<i>a</i>
diethylamine	328.6	346.7	18.1	NH	<i>b</i>
diisopropylamine	357.1	388.9	31.9	NH	<i>b</i>
dimethylamine	280.0	304.5	24.5	NH	<i>a</i>
diphenylamine	575.1	571.7	-3.5	NH	<i>b</i>
ethylamine	289.7	297.9	8.2	NH	<i>a</i>
ethyleneimine	329.0	351.4	22.4	NH	<i>a</i>
indole	526.1	534.0	7.9	NH	<i>a</i>
isobutylamine	340.9	348.9	8.1	NH	<i>b</i>
isopropylamine	305.6	319.4	13.8	NH	<i>b</i>
<i>m</i> -methoxyaniline	524.2	498.4	-25.8	NH	<i>a</i>
<i>m</i> -nitroaniline	557.2	548.7	-8.5	NH	<i>a</i>
methylamine	266.8	266.8	0.0	NH	<i>a</i>
<i>N</i> -allylaniline	492.2	503.5	11.3	NH	<i>a</i>
<i>n</i> -butylamine	350.6	353.8	3.2	NH	<i>b</i>
<i>n</i> -decylamine	493.7	483.7	-10.0	NH	<i>a</i>
<i>n</i> -dodecylamine	532.4	517.4	-15.0	NH	<i>a</i>
<i>N</i> -ethylbutylamine	381.2	392.3	11.1	NH	<i>a</i>
<i>n</i> -heptylamine	428.2	423.1	-5.1	NH	<i>a</i>
<i>n</i> -hexylamine	404.6	402.7	-2.0	NH	<i>b</i>

Table 2 (Continued)

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
<i>N</i> -methylbutylamine	364.2	374.6	10.4	NH	<i>a</i>
<i>N</i> -methylcyclohexylamine	422.2	425.2	3.0	NH	<i>a</i>
<i>N</i> -methylhexylamine	414.2	420.9	6.7	NH	<i>a</i>
<i>n</i> -nonylamine	475.4	462.8	-12.6	NH	<i>a</i>
<i>n</i> -octylamine	452.8	443.6	-9.2	NH	<i>a</i>
<i>n</i> -pentylamine	377.6	379.8	2.1	NH	<i>b</i>
<i>n</i> -propylamine	321.6	325.8	4.2	NH	<i>b</i>
<i>N</i> - <i>tert</i> -butylisopropylamine	371.2	402.5	31.3	NH	<i>a</i>
<i>o</i> -methoxyaniline	498.2	496.9	-1.3	NH	<i>a</i>
<i>p</i> -methoxyaniline	514.7	485.9	-28.8	NH	<i>a</i>
phenylhydrazine	516.7	527.0	10.3	NH	<i>a</i>
piperidine	379.6	398.7	19.1	NH	<i>b</i>
pyrrole	402.9	441.7	38.8	NH	<i>a</i>
pyrrolidine	359.7	376.0	16.3	NH	<i>a</i>
<i>sec</i> -butylamine	336.1	345.2	9.1	NH	<i>b</i>
<i>tert</i> -butylamine	317.6	337.4	19.8	NH	<i>b</i>
tri- <i>n</i> -butylamine	487.2	487.7	0.5	NH	<i>a</i>
triethylamine	516.2	543.5	27.3	NH	<i>a</i>
trimethylamine	276.0	276.8	0.8	NH	<i>b</i>
tripropylamine	429.7	429.5	-0.2	NH	<i>a</i>
1,3-dinitrobenzene	570.1	535.2	-35.0	N	Aldrich
1-methylpyrrole	386.0	374.6	-11.4	N	<i>a</i>
1-methylpyrrolidine	353.6	373.6	20.0	N	Aldrich
1-nitrobutane	425.7	401.9	-23.8	N	<i>a</i>
1-nitronaphthalene	577.1	555.1	-22.1	N	Aldrich
1-nitropropane	404.3	377.9	-26.4	N	<i>a</i>
2-methylpyridine	402.6	411.6	9.0	N	<i>b</i>
2-nitropropane	393.4	376.3	-17.1	N	<i>a</i>
9-methylcarbazole	616.8	566.0	-50.8	N	<i>a</i>
acetonitrile	354.8	323.9	-30.9	N	<i>a</i>
acridine	619.2	582.6	-36.6	N	<i>a</i>
acrylonitrile	350.5	365.5	15.0	N	<i>a</i>
adiponitrile	568.1	563.1	-5.0	N	<i>b</i>
<i>cis</i> -crotonitrile	380.6	400.1	19.5	N	<i>a</i>
glutaronitrile	559.1	537.5	-21.6	N	Aldrich
hexamethylene imine	404.9	418.6	13.7	N	<i>b</i>
hexanenitrile	436.8	451.0	14.2	N	<i>b</i>
hydrogen cyanide	298.8	277.4	-21.4	N	<i>a</i>
isobutyronitrile	376.8	392.7	16.0	N	<i>b</i>
<i>m</i> -nitrotoluene	503.7	482.8	-20.9	N	<i>a</i>
malononitrile	493.1	479.7	-13.5	N	Aldrich
<i>N,N</i> -dimethylaniline	466.7	439.6	-27.1	N	<i>b</i>
<i>n</i> -butyronitrile	390.8	394.9	4.2	N	<i>b</i>
nitrocyclopentane	453.2	437.2	-16.0	N	<i>a</i>
nitroethane	387.2	352.5	-34.7	N	<i>a</i>
nitromethane	374.4	324.7	-49.7	N	<i>a</i>
<i>o</i> -nitrotoluene	498.2	483.5	-14.7	N	<i>a</i>
<i>p</i> -nitrotoluene	511.7	484.2	-27.5	N	<i>a</i>
propionitrile	370.5	361.1	-9.4	N	<i>b</i>
pyrazine	388.6	410.8	22.2	N	Aldrich
pyridine	388.4	391.0	2.6	N	<i>a</i>
quinoline	510.8	498.3	-12.5	N	<i>b</i>
succinonitrile	539.6	511.6	-28.1	N	Aldrich
valeronitrile	414.4	423.9	9.4	N	<i>b</i>
1,1-diphenylethane	545.8	565.2	19.4	CH	<i>b</i>
1,2,3,4-tetrahydronaphthalene	480.8	476.4	-4.4	CH	<i>b</i>
1,2,3-trimethylbenzene	449.3	437.0	-12.3	CH	<i>b</i>
1,2,4-trimethylbenzene	442.5	438.7	-3.8	CH	<i>b</i>
1,3-butadiene	268.7	284.5	15.7	CH	<i>b</i>
1,3-cyclohexadiene	353.5	366.4	12.9	CH	<i>b</i>
1,5-hexadiene	332.6	351.6	19.0	CH	<i>b</i>
1-butene	266.9	278.3	11.4	CH	<i>b</i>
1-decene	443.8	440.7	-3.0	CH	<i>b</i>
1-dodecene	486.5	482.8	-3.7	CH	<i>b</i>
1-heptene	366.8	369.3	2.5	CH	<i>b</i>
1-hexadecene	558.0	556.8	-1.3	CH	<i>b</i>
1-octadecene	588.0	589.9	2.0	CH	<i>b</i>
1-octene	394.4	396.6	2.1	CH	<i>b</i>
1-tetradecene	524.3	524.2	-0.1	CH	<i>b</i>
2,2,3,3-tetramethylpentane	413.4	397.6	-15.8	CH	<i>b</i>
2,2,3-trimethylbutane	354.0	353.9	-0.2	CH	<i>b</i>
2,2,3-trimethylpentane	383.0	378.3	-4.7	CH	<i>b</i>
2,2,4-trimethylpentane	372.4	378.1	5.7	CH	<i>b</i>
2,2-dimethylbutane	322.9	328.6	5.7	CH	<i>b</i>
2,3,3-trimethylpentane	387.9	378.0	-9.9	CH	<i>b</i>
2,3-dimethyl-1-butene	328.8	338.0	9.2	CH	<i>b</i>
2,3-dimethyl-2-butene	346.4	337.7	-8.6	CH	<i>b</i>
2,3-dimethyl-3-butadiene	341.9	344.7	2.8	CH	<i>b</i>
2,3-dimethylbutane	331.1	329.7	-1.4	CH	<i>b</i>

Table 2 (Continued)

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
2,3-dimethylhexane	388.8	382.5	-6.3	CH	<i>b</i>
2,3-dimethylpentane	362.9	355.9	-7.1	CH	<i>b</i>
2,4,4-trimethyl-1-pentene	374.6	386.9	12.3	CH	<i>b</i>
2,4,4-trimethyl-2-pentene	378.1	386.7	8.6	CH	<i>b</i>
2-ethyl-1-butene	337.8	339.4	1.6	CH	<i>b</i>
2-methylpentane	333.4	332.0	-1.4	CH	<i>b</i>
2-methyl-1-butene	304.3	310.1	5.8	CH	<i>b</i>
2-methyl-1-pentene	335.0	339.7	4.7	CH	<i>a</i>
2-methyl-2-butene	311.7	310.0	-1.7	CH	<i>b</i>
2-methyl-2-pentene	340.4	340.2	-0.2	CH	<i>b</i>
2-methyl-3-ethylpentane	388.8	381.1	-7.7	CH	<i>b</i>
2-methylhexane	363.2	359.6	-3.6	CH	<i>b</i>
3,3-diethylpentane	419.3	400.2	-19.1	CH	<i>b</i>
3,3-dimethyl-1-butene	314.4	337.5	23.1	CH	<i>b</i>
3-methylpentane	336.4	330.6	-5.8	CH	<i>b</i>
3-methyl-1-butene	293.2	310.2	17.0	CH	<i>b</i>
3-methyl-1-pentene	327.3	340.5	13.1	CH	<i>b</i>
3-methylhexane	365.0	357.9	-7.1	CH	<i>b</i>
4-methyl-1-pentene	327.0	341.7	14.7	CH	<i>b</i>
$\alpha$ -methylstyrene	438.6	445.9	7.3	CH	<i>b</i>
benzene	353.2	372.6	19.4	CH	<i>b</i>
<i>cis</i> -1,2-dimethylcyclohexane	402.9	397.4	-5.5	CH	<i>b</i>
<i>cis</i> -1,3-dimethylcyclohexane	393.2	399.2	6.0	CH	<i>b</i>
<i>cis</i> -1,4-dimethylcyclohexane	397.5	399.0	1.5	CH	<i>b</i>
<i>cis</i> -2-butene	276.9	279.8	3.0	CH	<i>b</i>
<i>cis</i> -2-hexene	342.0	343.2	1.2	CH	<i>b</i>
cumene	425.6	440.6	15.1	CH	<i>b</i>
cyclohexane	353.9	351.9	-1.9	CH	<i>b</i>
cyclohexene	356.1	360.9	4.8	CH	<i>b</i>
cyclopentadiene	314.6	340.8	26.2	CH	<i>b</i>
cyclopentane	322.4	323.1	0.7	CH	<i>b</i>
cyclopentene	317.4	332.3	14.9	CH	<i>b</i>
diphenylmethane	537.4	551.2	13.8	CH	<i>b</i>
ethylbenzene	409.4	420.1	10.7	CH	<i>b</i>
ethylcyclohexane	404.9	400.2	-4.8	CH	<i>b</i>
ethylcyclopentane	376.6	376.4	-0.2	CH	<i>b</i>
1-hexene	336.6	343.6	7.0	CH	<i>b</i>
isobutane	261.4	269.2	7.8	CH	<i>b</i>
isobutene	266.3	266.7	0.5	CH	<i>b</i>
isobutylbenzene	445.9	461.6	15.7	CH	<i>b</i>
isopentane	301.0	302.2	1.2	CH	<i>b</i>
isoprene	307.2	316.4	9.2	CH	<i>b</i>
<i>m</i> -diethylbenzene	454.3	457.1	2.8	CH	<i>b</i>
<i>m</i> -diisopropylbenzene	476.3	498.0	21.7	CH	<i>b</i>
<i>m</i> -ethyltoluene	434.5	440.3	5.8	CH	<i>b</i>
<i>m</i> -toluidine	476.6	469.0	-7.6	CH	<i>b</i>
<i>m</i> -xylene	412.3	418.1	5.8	CH	<i>b</i>
mesitylene	437.9	438.4	0.6	CH	<i>b</i>
methylcyclohexane	374.1	375.8	1.7	CH	<i>b</i>
methylcyclopentadiene	345.9	367.5	21.6	CH	<i>b</i>
methylcyclopentane	345.0	351.5	6.5	CH	<i>b</i>
<i>n</i> -butane	272.6	270.7	-2.0	CH	<i>b</i>
<i>n</i> -butylbenzene	456.5	464.2	7.7	CH	<i>b</i>
<i>n</i> -butylcyclohexane	454.1	444.1	-10.0	CH	<i>b</i>
<i>n</i> -decane	447.3	432.4	-14.9	CH	<i>b</i>
<i>n</i> -dodecane	489.5	476.1	-13.4	CH	<i>b</i>
<i>n</i> -heptane	371.6	361.2	-10.4	CH	<i>b</i>
<i>n</i> -hexadecane	560.0	548.2	-11.8	CH	<i>b</i>
<i>n</i> -hexane	341.9	333.6	-8.3	CH	<i>b</i>
<i>n</i> -nonadecane	603.0	598.0	-5.0	CH	<i>b</i>
<i>n</i> -nonane	424.0	408.9	-15.1	CH	<i>b</i>
<i>n</i> -octadecane	589.9	579.1	-10.7	CH	<i>b</i>
<i>n</i> -octane	398.8	387.8	-11.0	CH	<i>b</i>
<i>n</i> -pentane	309.2	304.2	-5.0	CH	<i>b</i>
<i>n</i> -propylcyclohexane	429.9	423.5	-6.4	CH	<i>b</i>
<i>n</i> -propylbenzene	432.4	443.2	10.8	CH	<i>b</i>
<i>n</i> -propylcyclopentane	404.1	400.9	-3.2	CH	<i>b</i>
<i>n</i> -tetradecane	526.7	521.2	-5.6	CH	<i>b</i>
neopentane	282.6	299.9	17.3	CH	<i>b</i>
<i>o</i> -diethylbenzene	456.6	459.3	2.7	CH	<i>b</i>
<i>o</i> -ethyltoluene	438.3	438.9	0.6	CH	<i>b</i>
<i>o</i> -toluidine	473.6	455.9	-17.6	CH	<i>b</i>
<i>o</i> -xylene	417.6	417.1	-0.5	CH	<i>b</i>
<i>p</i> -cymene	450.3	460.5	10.2	CH	<i>b</i>
<i>p</i> -diethylbenzene	456.9	462.1	5.2	CH	<i>b</i>
<i>p</i> -diisopropylbenzene	483.6	498.6	14.9	CH	<i>b</i>
<i>p</i> -ethyltoluene	435.2	440.5	5.3	CH	<i>b</i>
<i>p</i> -toluidine	473.4	479.4	6.0	CH	<i>b</i>
<i>p</i> -xylene	411.5	419.0	7.4	CH	<i>b</i>

Table 2 (Continued)

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
1-pentene	303.1	312.7	9.6	CH	<i>b</i>
propane	231.1	231.2	0.1	CH	<i>b</i>
propylene	225.4	240.3	14.9	CH	<i>b</i>
<i>sec</i> -butylbenzene	446.8	461.1	14.2	CH	<i>b</i>
styrene	418.3	425.8	7.5	CH	<i>b</i>
toluene	383.8	396.3	12.5	CH	<i>b</i>
<i>trans</i> -1,2-dimethylcyclohexane	396.6	398.3	1.7	CH	<i>b</i>
<i>trans</i> -1,3-dimethylcyclohexane	397.6	398.1	0.5	CH	<i>b</i>
<i>trans</i> -1,4-dimethylcyclohexane	392.5	398.7	6.2	CH	<i>b</i>
<i>trans</i> -2-butene	274.0	279.7	5.7	CH	<i>b</i>
<i>trans</i> -2-hexene	341.0	343.0	1.9	CH	<i>b</i>
1,2-cyclohexanedione	467.1	478.6	11.4	O	Aldrich
1,2-propylene oxide	307.6	304.2	-3.4	O	<i>a</i>
1,3-dioxane	378.6	362.3	-16.4	O	Aldrich
1,4-dioxane	374.1	366.0	-8.1	O	Aldrich
1-butanal	348.0	353.7	5.7	O	<i>a</i>
1-heptanal	426.0	432.6	6.6	O	<i>a</i>
1-hexanal	401.4	408.7	7.3	O	<i>b</i>
1-pentanal	376.1	381.4	5.3	O	<i>a</i>
2,5-dihydrofuran	339.0	339.0	-0.0	O	<i>a</i>
2,5-hexanedione	464.1	455.7	-8.5	O	Aldrich
2-cyclohexen-1-one	441.1	422.3	-18.9	O	Aldrich
2-ethylhexyl acrylate	489.2	493.2	4.0	O	<i>a</i>
2-furaldehyde	435.1	425.2	-9.9	O	Aldrich
2-heptanone	424.0	422.5	-1.5	O	<i>a</i>
2-hexanone	400.9	398.5	-2.3	O	<i>b</i>
2-methylpropanal	337.3	342.9	5.6	O	<i>a</i>
2-methylfuran	337.6	366.2	28.5	O	Aldrich
2-nonanone	467.5	469.5	2.0	O	<i>a</i>
2-octanone	445.8	445.0	-0.8	O	<i>a</i>
2-pentanone	375.5	372.3	-3.2	O	<i>b</i>
3,3-dimethyl 2-butanone	379.4	388.8	9.4	O	<i>a</i>
3-hexanone	396.6	394.7	-2.0	O	<i>b</i>
3-methyl-2-pentanone	390.6	399.4	8.8	O	<i>a</i>
3-pentanone	375.1	362.0	-13.1	O	<i>a</i>
4-heptanone	417.2	411.6	-5.6	O	<i>a</i>
5-methyl-2-hexanone	418.0	419.4	1.4	O	<i>a</i>
5-nonanone	461.6	457.8	-3.8	O	<i>a</i>
acetaldehyde diethyl acetal	376.8	392.6	15.9	O	<i>b</i>
acetone	329.4	316.6	-12.8	O	<i>b</i>
acetophenone	475.1	471.3	-3.8	O	<i>b</i>
acetylacetone	413.6	427.6	14.1	O	<i>b</i>
acrylic aldehyde	325.8	323.3	-2.5	O	<i>b</i>
allyl acetate	377.1	377.6	0.4	O	<i>b</i>
anethole	508.5	485.6	-22.9	O	<i>a</i>
anthraquinone	653.1	644.1	-9.0	O	<i>a</i>
benzaldehyde	451.9	454.6	2.7	O	<i>b</i>
benzophenone	578.1	589.1	11.0	O	Aldrich
benzyl acetate	486.6	485.0	-1.7	O	<i>b</i>
benzyl benzoate	596.6	590.6	-6.0	O	<i>b</i>
benzyl ethyl ether	458.1	465.4	7.3	O	<i>a</i>
bicyclohexyl	512.2	495.8	-16.4	O	<i>b</i>
butyl vinyl ether	367.0	374.2	7.2	O	<i>b</i>
cycloheptanone	452.1	436.3	-15.9	O	Aldrich
cyclohexanone	428.9	416.3	-12.6	O	<i>b</i>
cyclopentanone	403.1	393.5	-9.6	O	Aldrich
di- <i>n</i> -butyl ether	413.4	413.8	0.4	O	<i>b</i>
di- <i>n</i> -butyl sulfoxide	523.1	500.9	-22.2	O	Aldrich
di- <i>n</i> -hexyl ether	498.9	496.8	-2.0	O	<i>b</i>
di- <i>n</i> -propyl ether	362.8	366.4	3.6	O	<i>b</i>
dibutyl phthalate	613.1	610.5	-2.7	O	<i>b</i>
dibutyl sebacate	622.1	627.8	5.7	O	<i>b</i>
diethyl ether	308.6	312.7	4.1	O	<i>b</i>
diethyl ketomalonate	482.1	488.5	6.3	O	Aldrich
diethyl maleate	498.2	477.0	-21.2	O	<i>a</i>
diethyl malonate	472.0	458.9	-13.1	O	<i>a</i>
diethyl oxalate	458.9	430.4	-28.5	O	<i>a</i>
diethyl phthalate	567.1	552.1	-15.0	O	<i>b</i>
di- <i>sec</i> -butyl ether	394.2	408.6	14.4	O	<i>a</i>
diisobutyl ketone	441.4	450.0	8.6	O	<i>a</i>
diisopropyl ether	341.4	363.1	21.7	O	<i>b</i>
diisopropyl ketone	397.6	407.5	9.9	O	<i>a</i>
dimethyl ether	248.3	243.7	-4.6	O	<i>a</i>
dimethyl phthalate	556.9	523.9	-33.0	O	<i>b</i>
dimethyl terephthalate	561.1	524.1	-37.0	O	<i>b</i>
diphenyl ether	531.5	539.8	8.4	O	<i>b</i>
divinyl ether	301.4	327.0	25.6	O	<i>b</i>
ethyl acetate	350.2	346.2	-4.0	O	<i>b</i>
ethyl acetoacetate	454.0	454.8	0.8	O	<i>a</i>

Table 2 (Continued)

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
ethyl acrylate	372.6	372.9	0.2	O	<i>b</i>
ethyl benzoate	486.6	478.3	-8.3	O	<i>b</i>
ethyl formate	327.5	314.0	-13.4	O	<i>b</i>
ethyl isobutyrate	383.0	390.6	7.6	O	<i>b</i>
ethyl isopropyl ether	326.1	339.3	13.2	O	<i>a</i>
ethyl isopropyl ketone	386.6	388.1	1.6	O	<i>b</i>
ethyl isovalerate	407.5	414.2	6.7	O	<i>a</i>
ethyl <i>n</i> -butyrate	394.6	392.7	-2.0	O	<i>b</i>
ethyl propionate	372.3	369.6	-2.6	O	<i>b</i>
ethyl propyl ether	337.0	341.2	4.2	O	<i>b</i>
ethyl vinyl ether	308.7	319.3	10.6	O	<i>b</i>
formaldehyde	254.0	236.4	-17.6	O	<i>a</i>
glutaraldehyde	460.0	453.9	-6.1	O	Aldrich
isobutyl acetate	389.8	392.0	2.2	O	<i>b</i>
isobutyl acrylate	405.1	416.4	11.2	O	<i>b</i>
isobutyl formate	371.2	364.6	-6.6	O	<i>b</i>
isobutyl isobutyrate	420.6	431.0	10.3	O	<i>b</i>
isopentyl isovalerate	467.2	473.6	6.4	O	<i>a</i>
isopropyl acetate	361.6	370.9	9.2	O	<i>b</i>
mesityl oxide	402.9	396.3	-6.6	O	<i>b</i>
methacrolein	341.1	348.1	7.0	O	<i>b</i>
methyl 2-furoate	454.1	430.1	-24.0	O	Aldrich
methyl acetate	330.1	317.5	-12.6	O	<i>b</i>
methyl acrylate	353.4	347.3	-6.1	O	<i>b</i>
methyl benzoylformate	520.1	522.9	2.8	O	Aldrich
methyl decanoate	505.0	493.4	-11.6	O	<i>a</i>
methyl ethyl ether	280.5	279.6	-0.9	O	<i>b</i>
methyl ethyl ketone	352.8	340.9	-11.9	O	<i>b</i>
methyl formate	304.9	282.5	-22.4	O	<i>a</i>
methyl isobutyl ether	331.7	338.2	6.5	O	<i>b</i>
methyl isobutyl ketone	389.6	399.0	9.3	O	<i>b</i>
methyl isopropyl ether	323.8	310.4	-13.3	O	<i>b</i>
methyl isopropyl ketone	367.6	367.8	0.3	O	<i>b</i>
methyl <i>n</i> -butyl ether	343.4	339.9	-3.5	O	<i>a</i>
methyl <i>n</i> -butyrate	375.9	369.1	-6.8	O	<i>a</i>
methyl <i>n</i> -pentyl ether	372.0	366.1	-5.9	O	<i>a</i>
methyl <i>n</i> -propyl ether	312.2	311.2	-1.0	O	<i>a</i>
methyl phenoxyacetate	516.1	495.3	-20.9	O	Aldrich
methyl propionate	352.6	344.4	-8.2	O	<i>b</i>
methyl pyruvate	408.6	401.7	-6.9	O	Aldrich
methyl <i>sec</i> -butyl ether	332.1	337.6	5.4	O	<i>b</i>
methyl <i>tert</i> -butyl ether	328.4	336.2	7.8	O	<i>b</i>
methyl <i>tert</i> -pentyl ether	359.4	361.2	1.7	O	<i>b</i>
methyl vinyl ether	278.6	288.0	9.3	O	<i>b</i>
methylal	315.0	315.1	0.1	O	<i>b</i>
<i>n</i> -butyl acetate	399.1	393.3	-5.9	O	<i>b</i>
<i>n</i> -butyl acrylate	421.0	418.4	-2.6	O	<i>b</i>
<i>n</i> -butyl ethyl ether	365.4	366.2	0.8	O	<i>b</i>
<i>n</i> -butyl formate	379.3	365.4	-13.9	O	<i>a</i>
<i>n</i> -butyl <i>n</i> -butyrate	438.2	437.2	-1.0	O	<i>a</i>
<i>n</i> -butyl propionate	419.8	415.7	-4.1	O	<i>a</i>
<i>n</i> -butyl stearate	623.1	663.0	39.8	O	<i>b</i>
<i>n</i> -decyl acetate	517.2	511.8	-5.4	O	<i>a</i>
<i>n</i> -heptyl acetate	465.6	457.9	-7.7	O	<i>a</i>
<i>n</i> -heptyl formate	451.3	439.9	-11.4	O	<i>a</i>
<i>n</i> -hexyl acetate	444.7	439.6	-5.1	O	<i>a</i>
<i>n</i> -hexyl formate	428.6	418.0	-10.6	O	<i>a</i>
<i>n</i> -octyl acetate	484.5	479.5	-5.0	O	<i>a</i>
<i>n</i> -octyl formate	472.0	459.2	-12.8	O	<i>a</i>
<i>n</i> -propionaldehyde	321.1	313.2	-8.0	O	<i>b</i>
<i>n</i> -propyl acetate	374.6	370.2	-4.5	O	<i>b</i>
<i>n</i> -propyl butyrate	416.5	415.1	-1.4	O	<i>a</i>
<i>n</i> -propyl propionate	395.6	393.3	-2.3	O	<i>b</i>
<i>p</i> -hydroquinone	558.1	550.9	-7.3	O	<i>b</i>
<i>p</i> -hydroxybenzaldehyde	583.2	538.5	-44.7	O	<i>a</i>
propionaldehyde diethyl acetal	395.9	412.8	16.9	O	Aldrich
salicylaldehyde	469.7	491.4	21.7	O	<i>a</i>
<i>sec</i> -butyl acetate	385.1	391.9	6.7	O	<i>b</i>
<i>sec</i> -butyl formate	366.5	364.6	-1.9	O	<i>a</i>
<i>tert</i> -butyl ethyl ether	346.0	362.2	16.2	O	<i>a</i>
<i>tert</i> -butyl acetate	369.1	391.2	22.0	O	<i>b</i>
<i>tert</i> -butyl formate	356.0	362.5	6.5	O	<i>a</i>
tetrahydrofuran	338.0	330.4	-7.6	O	<i>b</i>
tetrahydropyran	361.1	357.8	-3.4	O	Aldrich
<i>trans</i> -2,3-epoxybutane	327.6	333.6	6.0	O	Aldrich
valeraldehyde	376.1	379.1	3.0	O	<i>b</i>
1,2-benzenediol	518.7	558.7	40.0	OH	<i>a</i>
1,2-propylene glycol	460.8	465.9	5.2	OH	<i>b</i>
1,3-benzenediol	549.7	559.7	10.0	OH	<i>a</i>



Table 2 (Continued)

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
1,3-butanediol	480.1	480.4	0.3	OH	<i>b</i>
1,3-propylene glycol	487.6	475.5	-12.1	OH	<i>b</i>
1,4-butanediol	501.1	480.0	-21.2	OH	<i>b</i>
1,5-pentanediol	512.1	500.8	-11.3	OH	<i>b</i>
1,6-hexanediol	516.1	507.4	-8.7	OH	<i>b</i>
1-butanol	390.8	379.9	-10.9	OH	<i>a</i>
1-decanol	503.4	506.2	2.9	OH	<i>b</i>
1-heptanol	449.5	448.9	-0.6	OH	<i>a</i>
1-hexanol	430.1	429.3	-0.8	OH	<i>b</i>
1-methylcyclohexanol	441.2	448.8	7.6	OH	<i>a</i>
1-nonanoic acid	528.8	517.3	-11.5	OH	<i>a</i>
1-nonanol	486.3	487.8	1.5	OH	<i>a</i>
1-octanol	468.3	468.6	0.3	OH	<i>a</i>
1-pentanol	410.9	408.2	-2.7	OH	<i>b</i>
1-propanol	370.4	360.0	-10.4	OH	<i>a</i>
1-undecanol	518.2	523.5	5.3	OH	<i>a</i>
2,2-dimethyl-1-propanol	386.3	393.7	7.4	OH	<i>a</i>
2,3-butanediol	453.9	465.2	11.3	OH	<i>b</i>
2,3-xyleneol	490.1	483.5	-6.6	OH	<i>a</i>
2,4-xyleneol	484.1	492.3	8.2	OH	<i>a</i>
2,5-xyleneol	484.3	492.4	8.1	OH	<i>a</i>
2,6-dimethyl-4-heptanol	451.0	454.5	3.5	OH	<i>a</i>
2,6-xyleneol	474.2	479.5	5.2	OH	<i>b</i>
2-butanol	372.7	374.3	1.6	OH	<i>a</i>
2-ethylbutyric acid	466.9	458.9	-8.1	OH	<i>b</i>
2-ethyl-1-butanol	419.6	393.7	-25.9	OH	<i>b</i>
2-ethyl-1-hexanol	457.8	462.2	4.4	OH	<i>b</i>
2-heptanol	432.4	436.6	4.2	OH	<i>a</i>
2-hexanol	413.0	419.7	6.6	OH	<i>b</i>
2-methylbutyric acid	450.1	440.6	-9.5	OH	<i>b</i>
2-methyl-1-butanol	401.9	381.0	-20.8	OH	<i>b</i>
2-methyl-2-pentanol	395.0	401.3	6.3	OH	<i>b</i>
2-methyl-1-propanol	380.8	375.0	-5.8	OH	<i>a</i>
2-methyl-2-butanol	375.1	388.1	12.9	OH	<i>b</i>
2-nonanol	471.7	475.2	3.5	OH	<i>a</i>
2-octanol	453.0	457.8	4.8	OH	<i>a</i>
2-pentanol	392.1	396.9	4.8	OH	<i>b</i>
2-propanol	355.4	350.7	-4.7	OH	<i>b</i>
3,4-xyleneol	500.2	495.8	-4.4	OH	<i>a</i>
3,5-xyleneol	494.9	496.2	1.3	OH	<i>a</i>
3-methyl-1-butanol	404.4	397.7	-6.7	OH	<i>b</i>
3-methyl-2-butanol	384.6	395.0	10.4	OH	<i>b</i>
4-methyl-2-pentanol	404.9	411.2	6.4	OH	<i>b</i>
5-methyl-1-hexanol	445.2	446.2	1.0	OH	<i>a</i>
acetic acid	391.1	378.9	-12.2	OH	<i>a</i>
acrylic acid	412.1	394.8	-17.4	OH	Aldrich
allyl alcohol	370.2	370.3	0.0	OH	<i>b</i>
benzoic acid	522.4	513.1	-9.3	OH	<i>b</i>
benzyl alcohol	477.9	479.4	1.5	OH	<i>b</i>
cyclohexanol	434.0	435.0	1.0	OH	<i>b</i>
ethylene glycol	470.5	464.6	-5.9	OH	<i>a</i>
ethylenecarboxylic acid	414.1	397.3	-16.8	OH	<i>b</i>
formic acid	373.7	373.1	-0.6	OH	<i>a</i>
hexylene glycol	470.6	475.7	5.1	OH	<i>a</i>
isobutyric acid	427.9	415.9	-12.0	OH	<i>b</i>
isoquinoline	516.4	499.5	-16.9	OH	<i>a</i>
isovaleric acid	448.3	447.6	-0.7	OH	<i>b</i>
<i>m</i> -cresol	475.4	487.2	11.8	OH	<i>b</i>
methacrylic acid	434.2	426.7	-7.5	OH	<i>a</i>
methanol	337.8	314.7	-23.1	OH	<i>a</i>
<i>n</i> -butyric acid	436.4	408.6	-27.8	OH	<i>b</i>
<i>n</i> -decanoic acid	543.2	522.2	-21.0	OH	<i>a</i>
<i>n</i> -dodecanoic acid	571.9	561.3	-10.6	OH	<i>a</i>
<i>n</i> -heptanoic acid	496.2	483.7	-12.5	OH	<i>a</i>
<i>n</i> -hexanoic acid	478.9	459.4	-19.5	OH	<i>b</i>
<i>n</i> -tridecanoic acid	585.3	574.4	-10.9	OH	<i>a</i>
<i>n</i> -undecanoic acid	557.4	534.4	-23.0	OH	<i>a</i>
neopentanoic acid	437.0	438.0	1.0	OH	<i>a</i>
neopentyl glycol	483.0	484.0	1.0	OH	<i>b</i>
<i>o</i> -cresol	464.1	481.2	17.0	OH	<i>b</i>
<i>p</i> -cresol	475.1	481.2	6.1	OH	<i>b</i>
<i>p</i> -ethylphenol	491.1	498.6	7.5	OH	<i>a</i>
<i>p</i> - <i>tert</i> -butylphenol	512.9	527.6	14.7	OH	<i>a</i>
<i>p</i> - <i>tert</i> -octylphenol	563.6	592.0	28.4	OH	<i>a</i>
phenol	455.0	466.1	11.1	OH	<i>b</i>
propionic acid	414.3	409.2	-5.1	OH	<i>b</i>
stearic acid	648.4	639.7	-8.7	OH	<i>b</i>
<i>tert</i> -crotonic acid	458.1	433.4	-24.7	OH	<i>b</i>
valeric acid	458.6	440.5	-18.2	OH	<i>b</i>

Table 2 (Continued)

compound	BP <sub>exp</sub>	BP <sub>calc</sub>	BP <sub>exp</sub> - BP <sub>calc</sub>	subset	source
1,2-ethanedithiol	419.2	418.0	-1.2	S	a
2-methyl thiophene	385.7	401.0	15.3	S	a
tert-butyl disulfide	473.6	495.7	22.1	S	Aldrich
diethyl sulfite	481.1	534.0	52.9	S	Aldrich
diethyl sulfite	432.1	469.6	37.5	S	Aldrich
dimethyl disulfide	382.9	376.5	-6.4	S	a
dimethyl sulfate	461.1	501.8	40.6	S	Aldrich
dimethyl sulfide	310.5	288.1	-22.4	S	a
dimethyl sulfite	399.6	431.3	31.7	S	Aldrich
diphenyl sulfone	652.1	663.2	11.0	S	Aldrich
dipropyl sulfone	543.1	535.3	-7.9	S	Aldrich
ethyl mercaptan	308.2	307.4	-0.8	S	a
ethyl tert-butyl sulfide	393.6	395.4	1.8	S	a
isopropyl mercaptan	325.7	336.2	10.5	S	a
methyl mercaptan	279.1	275.6	-3.5	S	a
methyl n-butyl sulfide	396.6	375.5	-21.1	S	a
methyl n-propyl sulfide	368.7	349.8	-18.9	S	a
methyl tert-butyl sulfide	372.0	370.5	-1.5	S	a
n-butyl mercaptan	371.6	368.0	-3.6	S	a
n-decyl mercaptan	512.3	508.6	-3.7	S	a
n-heptyl mercaptan	450.1	450.2	0.1	S	a
n-hexyl mercaptan	425.8	426.1	0.3	S	a
n-nonyl mercaptan	493.0	495.0	2.0	S	a
n-octyl mercaptan	472.2	472.5	0.3	S	a
n-pentyl mercaptan	399.8	401.7	1.9	S	a
n-propyl mercaptan	340.9	341.8	0.9	S	a
sec-butyl mercaptan	358.1	363.0	4.9	S	a
sulfolane	558.1	514.8	-43.4	S	Aldrich
tert-butyl mercaptan	337.4	359.6	22.2	S	a
tetrahydrothiophene	394.3	366.0	-28.3	S	a
thiophene	357.3	379.2	21.9	S	a

<sup>a</sup> Reference 6. <sup>b</sup> Reference 26.

chemically calculated atomic charges were also used to calculate the charged partial surface area (CPSA) descriptors introduced by Stanton and Jurs.<sup>20,21</sup> Altogether more than 800 molecular descriptors were calculated. The actual number of descriptors calculated for each compound depends on the structural diversity of the compound. Descriptors, which were not defined for every compound (e.g. partial charge on a Cl atom), were either omitted or padded with zeros in the subsequent development of QSPR correlation equations.

The heuristic procedure available in the CODESSA program was used to develop multilinear correlation models from the pool of calculated descriptors. This procedure (described in detail elsewhere<sup>1</sup>) is based on the stepwise regression technique<sup>22</sup> that proceeds from the statistical significance and collinearity control of the descriptors selected into the correlation equations.

**Correlation Analysis of Subsets.** First, the original set of 612 compounds was divided into 9 subsets according to the elements represented in a compound: (i) set **CH** included 113 compounds containing only C and H atoms; (ii) set **O** included 138 oxygen containing compounds without hydrogen bonding; (iii) set **OH** included 85 oxygen containing compounds capable of hydrogen bonding; (iv) set **N** included 39 nitrogen containing compounds without hydrogen bonding; (v) set **NH** included 59 nitrogen containing compounds capable of hydrogen bonding; (vi) set **F** included 30 fluorine containing compounds; (vii) set **Cl** included 54 chlorine containing compounds; (viii) set **BrI** included 34 compounds containing bromine or iodine; and (ix) set **S** consisted of 32 sulfur containing compounds. The remaining 28 compounds

**Table 3.** Correlation Analysis of Subsets of Compounds (*n*, Number of Structures; *m*, Number of Descriptors in a Model; *R*<sup>2</sup> and *s*, Correlation Coefficient and Standard Deviation of a Correlation Equation, Respectively)

subset	<i>n</i>	<i>R</i> <sup>2 a</sup>	<i>s</i> <sup>a</sup>	<i>R</i> <sup>2 b</sup>	<i>s</i> <sup>b</sup>	<i>R</i> <sup>2 c</sup>	<i>s</i> <sup>c</sup>	<i>m</i> <sup>c</sup>
CH	113	0.9908	7.9	0.9908 <sup>d</sup>	7.9	0.9908	7.6	1
O	138	0.8934	25.4	0.8934 <sup>d</sup>	25.4	0.9772	11.9	3
OH	85	0.7177	31.7	0.9224	16.7	0.9463	14.0	3
N	39	0.6421	52.1	0.6421 <sup>d</sup>	52.1	0.9617	17.5	3
NH	59	0.8921	26.7	0.9607	16.2	0.9607	16.2	2
S	32	0.8201	35.6	0.8714	30.6	0.9616	17.0	3
F	30	0.5322	59.6	0.5322 <sup>d</sup>	59.6	0.9496	20.3	3
Cl	54	0.8576	23.5	0.8576 <sup>d</sup>	23.5	0.9291	16.9	3
Br, I	34	0.9716	12.3	0.9716 <sup>d</sup>	12.3	0.9716	12.3	1

<sup>a</sup> One-parameter correlation equation with the cubic root of the gravitation index  $G_1^{1/3}$ . <sup>b</sup> Correlation equation with the cubic root of the gravitation index  $G_1^{1/3}$  and HDSA2. <sup>c</sup> Best correlation equation for the subset. <sup>d</sup> Effectively the one-parameter equation with the cubic root of the gravitation index because HDCA2 values are zeroes for all compounds in the subset.

did not fit into any specific subset and were used to test the correlation model (see Table 6).

For each subset, our previously reported two-parameter correlation model (involving  $G_1^{1/3}$  and HDSA2)<sup>1</sup> was refitted. Because six out of the nine subsets did not contain compounds capable of hydrogen bonding, the second parameter (HDSA2) had zero values for all compounds in these subsets. Therefore, one-parameter fits were calculated for each subset using only the cubic root of the gravitation index  $G_1^{1/3}$  (column 3 in Table 3) [We also found that the square root of the gravitation index performs as well as the cube root, and in all our regressions the cube root of the gravitation index can be replaced by the square root without any loss

of fitness.] Obtained fits were very poor for **N** and **F** subsets, better for **O**, **OH**, **NH**, **S**, and **Cl** subsets and good for **CH** and **BrI** subsets.

The addition of the hydrogen donors charged surface area descriptor HDSA2 improved the correlations for subsets **OH**, **NH**, and **S** as expected (columns 5 and 6 in Table 3). These regression results demonstrate, however, that the previously reported two-parameter equation<sup>1</sup> does not model the **F** containing compounds and also needs substantial improvement for the **N**, **S**, and **Cl** containing compounds. In the case of the **F** containing compounds, the errors in the prediction of normal boiling points were largest for poly-fluorinated compounds. Balaban *et al.* have noted<sup>23</sup> that, in contrast with other halogenated hydrocarbons, the fluorinated hydrocarbons show a parabolic dependence of their boiling point on the extent of fluorination; i.e., the boiling point first increases with a replacement of one, two, etc., hydrogens with fluorine, but then it starts to decrease when more hydrogens are replaced by fluorine. Among the nitrogen containing compounds, the largest errors of prediction were found for nitriles. In the case of the subset of the sulfur containing compounds, the sulfates and sulfites had the largest deviations from the experimental values.

For the **CH** subset of compounds, the cube root of the gravitation index  $G_1^{1/3}$  alone gave an excellent correlation ( $R^2 = 0.9923$ ). The boiling points of aliphatic, nonaliphatic, cyclic, acyclic, and aromatic hydrocarbons can be predicted with a standard error of less than 8 K using the following equation:

$$T_b = -190.7 + 68.6G_1^{1/3} \quad (3)$$

$$n = 113, \quad R^2 = 0.9908, \quad F = 11965, \quad s = 7.6$$

For the **BrI** subset the cube root of the gravitation index also provided a good model:

$$T_b = -172.8 + 66.6G_1^{1/3} \quad (4)$$

$$n = 34, \quad R^2 = 0.9761, \quad F = 1094, \quad s = 12.3$$

For the **NH** subset of compounds, the combination of the cube root of the gravitation index and the hydrogen bonding descriptor HDSA2 was successful in producing a good predictive model:

$$T_b = -253.6 + 72.9G_1^{1/3} + 24134\text{HDSA2} \quad (5)$$

$$n = 59, \quad R^2 = 0.9607, \quad F = 684, \quad s = 16.2$$

For the remaining six subsets, the correlation analysis was performed with the inclusion of all available descriptors to determine which parameters are successful in capturing structural features essential for determining their boiling points. A summary of the best correlations obtained for different subsets of organic compounds is given in Table 3 (columns 7–9).

For the **O** subset, the best correlation model includes the cube root of the gravitation index, the surface area of hydrogen acceptors in the molecule, HASA1, and the

topographic electronic index  $T_1$ :

$$T_b = -215 + 74.1G_1^{1/3} + 0.75\text{HASA1} - 57.3T_1 \quad (6)$$

$$n = 138, \quad R^2 = 0.9772, \quad F = 1914, \quad s = 11.9$$

The HASA1 descriptor is calculated as a sum of exposed surface areas of all possible hydrogen acceptor atoms in the molecule:

$$\text{HASA1} = \sum_A S_A \quad (7)$$

The following atoms were considered as possible hydrogen acceptors: carbonyl oxygen atoms (except in COOR), hydroxy oxygen atoms, amino nitrogen atoms, aromatic nitrogens, and mercapto sulfur atoms. The topographic electronic index is calculated as

$$T_1 = \sum_{i < j}^{N_B} \frac{|q_i - q_j|}{r_{ij}^2} \quad (8)$$

where  $q_i$  denotes a partial charge on the  $i$ th atom,  $r_{ij}$  denotes a distance between  $i$ th and  $j$ th atoms, and summation is done over all bonded pairs of atoms in the molecule. The atomic partial charges used in this formula were calculated using an empirical method proposed by Zefirov *et al.*<sup>24</sup>

The best correlation model for the **OH** subset includes, in addition to the cube root of the gravitation index and the hydrogen donors charged surface area, also the above described hydrogen acceptors surface area descriptor, HASA1:

$$T_b = -131 + 60.7G_1^{1/3} + 13945\text{HDSA2} + 0.65\text{HASA1} \quad (9)$$

$$n = 86, \quad R^2 = 0.9463, \quad F = 481, \quad s = 14.0$$

As mentioned above, for the **N** subset of compounds the largest prediction errors by the one-parameter correlation equation with the gravitation index were observed for the nitriles. Interestingly, the most significant improvement of the correlation model was achieved by using the simple count of nitrile groups. The comparison of the total dipole moments calculated by the AMPAC with experimental values revealed that the AM1 calculated dipole moments were systematically lower for nitriles. Also, the Mulliken partial charges on the carbon and nitrogen atoms of a nitrilo group were unexpectedly small. Apparently, the AM1 model is not parametrized well for the description of the charge distribution in this class of compounds, and, therefore, it is inappropriate to use the charge related descriptors obtained for other classes of compounds for the QSPR description of the boiling points of nitriles. The final best correlation model for the **N** subset of compounds included also the topographic electronic index:

$$T_b = -204.5 + 76G_1^{1/3} + 76.7N_{\text{CN}} - 95.4T_1 \quad (10)$$

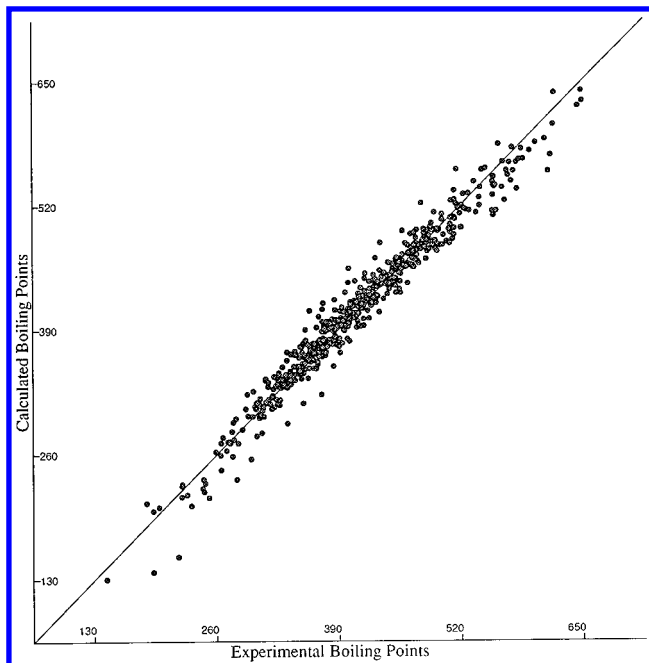
$$n = 39, \quad R^2 = 0.9617, \quad F = 293, \quad s = 17.5$$

The two-parameter model for the **S** subset was significantly improved by the addition of the maximum nuclear—

**Table 4.** Details of the Best Correlation Model for the Boiling Points of 584 Diverse Organic Compounds

descriptor	$X$	$\Delta X$	$t$ -test	$R^2$ <sup>a</sup>	$s$ <sup>a</sup>
$G_I^{1/3}$	64.6	0.73	88.2	0.6725	46.9
HDSA2	536	523.2	33.5	0.7819	38.3
$N_F/N$	-193.0	-10.8	17.8	0.8782	28.6
$N_{CN}$	-86.0	3.01	28.5	0.9192	23.3
HASA1	0.75	0.037	20.1	0.9371	20.6
$T_I$	-85.8	4.69	18.2	0.9469	18.9
CSA2 <sub>H</sub>	10.4	0.68	15.2	0.9574	17.0
CSA2 <sub>Cl</sub>	21.9	2.06	10.7	0.9645	15.5
intercept	-166.5	5.3	31.4		

<sup>a</sup> Column shows the improvement of the correlation from the top of the table to the bottom with an addition of the next descriptor.

**Figure 1.** Plot of predicted versus experimental boiling points for the set of 584 organic compounds.

nuclear repulsion term for a C–O bond ( $NNR_{CO,max}$ ):

$$T_b = -156.3 + 66.2G_I^{1/3} + 63107HDSA2 - 0.40NNR_{CO,max} \quad (11)$$

$$n = 32, R^2 = 0.9616, F = 233, s = 17.0$$

This descriptor has the highest values for sulfates and sulfites and helps to improve the performance of the model for these compounds. However, the appearance of this descriptor in the QSPR model is physically unclear and may be due to a statistical effect.

The **F** subset gave the poorest correlation with the cube root of the gravitation index. It is well-known that perflu-

orinated compounds have small cohesion energies and low boiling points. Therefore, it is not surprising that the gravitation index reflecting the dispersion interaction in hydrocarbons does not perform well for fluorocarbons. The descriptor accounting for most of the remaining variation in boiling points turned out to be a simple count of fluorine atoms normalized by the total number of atoms in the molecule (relative number of F atoms,  $N_F/N$ ). Some of the compounds in the F subsets also had nitrilo groups, and addition of the count of nitrilo groups led to a satisfactory three-parameter model:

$$T_b = 20.1 + 42.7G_I^{1/3} - 275.9N_F/N + 61.5N_{CN} \quad (12)$$

$$n = 30, R^2 = 0.9496, F = 163, s = 20.3$$

For the **Cl** subset the best correlation model involved the gravitation index, the maximum partial charge for a Cl atom, and the fractional partial negatively charged surface area (FNSA3) descriptor:<sup>20</sup>

$$T_b = -231.4 + 61.3G_I^{1/3} - 31.0Q_{Cl,max} - 999.3FNSA3 \quad (13)$$

$$n = 54, R^2 = 0.9291, F = 218, s = 16.9$$

Both the maximum partial charge for a Cl atom and the FNSA3 descriptor were calculated using atomic partial charges obtained by the empirical method proposed by Zefirov *et al.*<sup>24</sup> Analogous descriptors based on the Mulliken partial charges derived from AMPAC calculations did not perform quite as well.

**Correlation Analysis of the Combined Set.** Nine descriptors involved in the reported correlation models for the subsets were combined and regressed against the combined set of 584 compounds. The resulting correlation equation had correlation coefficient  $R^2 = 0.9517$  and standard error  $s = 18.1$  K. However, three descriptors, the maximum nuclear–nuclear repulsion term for a C–O bond, the maximum partial charge for a Cl atom, and the fractional partial negatively charged surface area FNSA3, were not significant when applied to the whole set. The elimination of these descriptors led to a six-parameter correlation model with the correlation coefficient  $R^2 = 0.9469$  and standard error  $s = 18.9$  K: All six remaining descriptors were statistically significant based on the  $t$ -test. After this procedure, all other available descriptors were tested for whether or not they improved significantly the model. The most significant improvement was obtained after inclusion of two further descriptors, the charged surface area of hydrogen atoms (CSA2<sub>H</sub>) and the charged surface area of

**Table 5.** Correlation Matrix (R) of the Descriptors Involved in the Best Correlation Model for the Boiling Points of 584 Diverse Organic Compounds

	$G_I^{1/3}$	HDSA2	$N_{CN}$	$N_F/N$	HASA1	$T_I$	CSA2 <sub>H</sub>	CSA2 <sub>Cl</sub>
$G_I^{1/3}$	1.0000	-0.1377	-0.1134	0.0487	-0.0789	0.5419	0.5612	-0.0926
HDSA2	-0.1377	1.0000	-0.0834	-0.0999	0.4658	0.3727	0.1162	-0.1384
$N_{CN}$	-0.1134	-0.0834	1.0000	0.0281	-0.1033	-0.1497	-0.1842	-0.0282
$N_F/N$	0.0487	-0.0999	0.0281	1.0000	-0.0924	0.0891	-0.3329	-0.0531
HASA1	-0.0789	0.4658	-0.1033	-0.0924	1.0000	0.1806	0.0186	-0.1277
$T_I$	0.5419	0.3727	0.0891	0.0891	0.1806	1.0000	0.7167	0.2721
CSA2 <sub>H</sub>	0.5612	0.1162	-0.1842	-0.3329	0.0186	0.7167	1.0000	-0.2734
CSA2 <sub>Cl</sub>	-0.0926	-0.1384	-0.0282	-0.0531	-0.1277	0.2721	-0.2734	1.0000

**Table 6.** Test Prediction of Boiling Points of Polyfunctional Compounds Using the Best Correlation Equation Developed for the Set of 584 Organic Compounds

compound	expt $T_b$	calc $T_b$	deviation
1,1-dichloro-2,2,2-trifluoroethane	301.00	293.24	-7.76
1,1-dichlorotetrafluoroethane	276.20	275.31	-0.89
1,2-dibromotetrafluoroethane	320.15	325.51	5.36
1,2-dichlorotetrafluoroethane	276.95	285.62	8.67
1,3-dichloro-4-iodobenzene	535.65	520.84	-14.81
1,3-dichlorotetrafluorobenzene	424.15	429.62	5.47
1,4-dibromo-2-fluorobenzene	489.15	493.78	4.63
1-bromo-2,4,5-trifluorobenzene	417.15	428.50	11.35
1-bromo-2,4,6-trifluorobenzene	413.65	428.64	14.99
1-bromo-4-fluorobenzene	423.15	440.89	17.74
1-chloro-2-iodobenzene	507.65	496.09	-11.56
1-chloro-4-iodobenzene	499.65	497.36	-2.29
2,3-dichlorohexafluoro-2-butene	340.15	366.77	26.62
$\alpha$ -bromo-2,3,4,5,6-pentafluorotoluene	447.65	444.21	-3.44
bromopentafluorobenzene	410.15	406.59	-3.56
bromotrichloromethane	378.00	378.77	0.77
chlorodifluoromethane	232.30	222.51	-9.79
chloropentafluoroethane	234.15	239.23	5.08
chlorotrifluoromethane	191.70	185.55	-6.15
dichlorofluoromethane	282.00	282.95	0.95
<i>m</i> -chloroaniline	501.70	498.99	-2.71
<i>m</i> -chlorophenol	487.00	496.81	9.81
<i>o</i> -bromoaniline	502.20	514.75	12.55
<i>o</i> -fluoroaniline	455.70	463.91	8.21
<i>p</i> -chloroaniline	503.70	496.30	-7.40
<i>p</i> -chlorophenol	493.10	495.79	2.69
<i>p</i> -fluorobenzylamine	456.20	462.55	6.35
4-hydroxy-3-methoxybenzaldehyde	558.00	565.60	7.60

chlorine atoms (CSA2<sub>C</sub>) calculated as follows:

$$\text{CSA2} = \sum_i q_i S_i^{1/2} \quad (14)$$

where  $q_i$  denotes the Mulliken partial charge on the  $i$ th atom and  $S_i$  denotes the exposed surface area of the  $i$ th atom in the compound. Further attempts to improve the correlation model were not successful. The final best eight-parameter correlation model obtained for the set of 584 diverse organic compounds had correlation coefficient  $R^2 = 0.9645$  and standard error  $s = 15.5$  K. The details of this correlation equation are given in Table 4. The corresponding graph of predicted boiling points versus experimental boiling points is given in Figure 1. The correlation matrix (Table 5) demonstrates that the descriptors involved in the model for predicting boiling points are almost orthogonal except for the topographic index and charged surface area of hydrogen atoms. These descriptors have an intercorrelation coefficient  $R^2 = 0.5136$  which is still an acceptable value.<sup>25</sup> Predicted boiling point values and prediction errors are given in Table 2; the calculated root-mean-square error for the entire set of 584 boiling points is 14.6 K.

The best correlation model was used to predict boiling points for a set of 28 additional organic compounds containing several different heteroatoms. A comparison of predicted and experimental boiling points is given in Table 6. Most of the predicted values are well within the standard error except for the 2,3-dichlorohexafluoro-2-butene, 1,3-dichloro-4-iodobenzene, and 1-bromo-4-fluorobenzene deviations which are 23.5, 20.3, and 17.7 K, respectively. Notably, most of these 28 compounds are from the fluorinated and chlorinated compounds which had the highest standard errors among the respective subsets.

**Table 7.** Test Prediction of Boiling Points of Simple Inorganic Compounds Using the Best Correlation Equation Developed for the Set of 584 Organic Compounds

compound	expt $T_b$	calc $T_b$	deviation
CH <sub>3</sub> Cl	249.0	238.9	10.0
CH <sub>3</sub> F	195.0	157.3	37.6
CH <sub>3</sub> NH <sub>2</sub>	267.0	266.9	0.1
CH <sub>3</sub> OH	338.0	318.6	19.3
CH <sub>3</sub> SH	267.0	275.6	8.7
CH <sub>4</sub>	112.0	109.8	2.2
H <sub>2</sub> O	373.0	435.0	62.0
H <sub>2</sub> O <sub>2</sub>	425.0	544.1	119.1
H <sub>2</sub> S	213.0	199.7	13.2
H <sub>2</sub> S <sub>2</sub>	344.0	392.9	48.9
HBr	206.0	167.3	38.6
HCl	188.0	246.2	58.2
HF	293.0	303.2	10.2
N <sub>2</sub> H <sub>4</sub>	387.0	378.4	8.6
NH <sub>2</sub> OH	329.5	352.8	23.3
NH <sub>3</sub>	240.0	215.2	24.8
P <sub>2</sub> H <sub>4</sub>	330.5	283.9	46.4
PH <sub>3</sub>	185.0	119.0	66.0
SiH <sub>4</sub>	161.0	131.6	29.4

In our previous paper on boiling point correlation we found that the boiling points of a range of simple inorganic compounds could be predicted reasonably well by our correlation equation developed for organic compounds.<sup>1</sup> This increased our confidence that the equations developed do reflect the intrinsic physicochemical properties which determine the boiling point of a compound. Therefore we applied the extended model reported here to predict the boiling points of 19 simple inorganic compounds. Predicted and experimental boiling points are presented in Table 7: for most compounds deviations are less than 50 K with the highest errors for hydrogen peroxide (119 K), phosphine (66 K), and water (62 K).

## CONCLUSIONS

Successful correlation equations were developed for the normal boiling points of different subsets of organic compounds containing various heteroatoms. These equations are based on our previously reported<sup>1</sup> two-parameter model which captures the most significant variations in boiling point. The model has been extended by adding descriptors specific for the subsets. The resulting individual QSPR correlation equations involve one to three parameters and have standard errors ranging from 8 K for hydrocarbons to 17 K for chlorinated compounds and 20 K for fluorinated compounds.

Proceeding from the correlation equations for the subsets of compounds, a general eight-parameter correlation model was developed for the prediction of the boiling point of any organic compound containing C, H, O, N, S, F, Cl, Br, and I atoms. This correlation model covers a larger diversity of organic structures than any other predictive equation obtained by regression analysis techniques (cf. Table 1) and offers a standard prediction error of 15.5 K. For comparison, the experimental error of the boiling points from the DIPPR database has been assessed as high as 11.4 K<sup>6,26</sup> and is even higher for the boiling points from the Aldrich catalog. In conclusion, the QSPR equations developed in the present study enable the confident prediction of the normal boiling points of organic compounds on the basis of their chemical

structure alone. The normal boiling point can thus be predicted for any organic compound, including those not yet synthesized or otherwise unavailable, which makes these presented equations attractive for many technological, environmental, and chemical applications.

#### ACKNOWLEDGMENT

We thank Mr. Shaobin Du for help in collecting data and in the calculations, and Ms. Hong Wu for technical assistance.

#### REFERENCES AND NOTES

- (1) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. 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.
- (2) Fisher, C. H. Boiling Point Gives Critical Temperature. *Chem. Eng.* **1989**, *96*, 157–158.
- (3) Satyanarayana, K.; Kakati, M. C. Note: Correlation of Flash Points. *Fire Mater.* **1991**, *15*, 97–100.
- (4) 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.
- (5) Horvath, A. L. In *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds*; Elsevier: Amsterdam, 1992.
- (6) Wessel, M. D.; Jurs, P. C. 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.
- (7) Le, T. D.; Weers, J. G. QSPR and GCA Models for Predicting the Normal Boiling Points of Fluorocarbons. *J. Phys. Chem.* **1995**, *99*, 6739–6747.
- (8) (a) Stein, S. E.; Brown, R. L. Estimation of Normal Boiling Points from Group Contributions. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 581–587. (b) Wang, S.; Milne, G. W. A.; Klopman, G. Graphic Theory and Group Contributions in the Estimation of Boiling Points. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1242–1250.
- (9) Karelson, M. Quantum-Chemical Treatment of Molecules in Condensed Disordered Media. *Adv. Quantum Chem.* **1997**, *28*, 141–157.
- (10) Weast, R. C. *Handbook of Chemistry and Physics*; CRC Press: Cleveland, OH, 1976.
- (11) *Aldrich Catalog Handbook of Fine Chemicals*; Aldrich: Milwaukee, WI, 1996.
- (12) *PCMODEL*, 5th ed., Serena Software: Bloomington, IN, 1992.
- (13) *AMPAC 5.0*; Semichem: Shawnee, KS, 1994.
- (14) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (15) *CODESSA 2.0*; University of Florida: Gainesville, FL, 1994.
- (16) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, *24*, 279–287.
- (17) Stewart J. J. P. *MOPAC Program Package. QCPE* **1989**, 455.
- (18) Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Quantitative Structure–Property Relationship Treatment. *Anal. Chem.* **1994**, *66*, 1799–1807.
- (19) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027–1043.
- (20) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure–Property Relationship Studies. *Anal. Chem.* **1990**, *62*, 2323–2329.
- (21) Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Pyrans and Pyrroles. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306–316.
- (22) Draper, N. R.; Smith, H. *Applied Regression Analysis*; Wiley: New York, 1966.
- (23) Balaban, A. T.; Joshi, N.; Kier, L. B.; Hall, L. H. Correlation between Chemical Structure and Normal Boiling Points of Halogenated Alkanes C<sub>1</sub>–C<sub>4</sub>. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 233–237.
- (24) Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I. Scheme for the Calculation of the Electronegativities of Atoms in a Molecule in the Framework of Sanderson's Principle. *Dokl. Akad. Nauk SSSR (Engl. Transl.)* **1987**, *296*, 440–443.
- (25) *Reading and Understanding Multivariate Statistics*; Grimm, L. G., Yarnold, P. R., Eds.; American Psychological Association: Washington, DC, 1994; p 45.
- (26) Egolf, L. M.; Wessel, M. D.; Jurs, P. C. Prediction of Boiling Points and Critical Temperatures of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 947–956.
- (27) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186–4194.
- (28) Wessel, M. D.; Jurs, P. C. Prediction of Normal Boiling Points of Hydrocarbons from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 68–76.
- (29) Stanton, D. T.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Furans, Tetrahydrofurans, and Thiophenes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301–310.
- (30) Egolf, L. M.; Jurs, P. C. Prediction of Boiling Points of Organic Heterocyclic Compounds Using Regression and Neural Network Techniques. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 616–625.
- (31) Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. D. Correlation between Structure and Normal Boiling Points of Haloalkanes C<sub>1</sub>–C<sub>4</sub> Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118–1121.
- (32) Balaban, A. T.; Kier, L. B.; Joshi, N. 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.
- (33) Simamora, P.; Miller, A. H.; Yalkowsky, S. H. Melting Point and Normal Boiling Point Correlations: Applications to Rigid Aromatic Compounds. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 437–440.

CI970029V