

Relationships of Critical Temperatures to Calculated Molecular Properties

Alan R. Katritzky* and Lan Mu

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

Mati Karelson

Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu, EE 2400, Estonia

Received August 30, 1997

Quantitative structure–property relationships (QSPR) of critical temperatures with small numbers of physically significant molecular descriptors are developed using the CODESSA (comprehensive descriptors for structural and statistical analysis) technique. A highly significant one-parameter model correlates to the critical temperatures of 76 hydrocarbons at $R^2 = 0.953$. A successful three-parameter model for 165 diverse compounds ($R^2 = 0.955$) reveals fundamental structural influences on liquid-state properties.

INTRODUCTION

The critical temperature is one of the important properties in revealing the intermolecular interactions between the molecules in liquid state. As established by the principle of corresponding states, all substances obey the same equation of state in terms of reduced variables, the temperature, the pressure, and the volume scaled by the critical temperature, critical pressure, and critical volume. The reduced variables characterize a specific state of a substance. Therefore, the scaling factors that are the critical properties express the individual features of each substance and are determined by the molecular structural differences between the substances.

Efforts to establish models for the calculation of critical temperatures can be traced back to early engineers' estimates of critical temperatures via measured quantities such as boiling points, Parachor, and molar refraction.¹ More direct quantitative structure–property relationships (QSPR) as applied in the chemical engineering community were expressed by the original group contribution method² and its modified approach,³ which predefine for each structural group a constant increment to a property such as critical temperature (note: each property requires a different set of constants). Attempts to link the decisive structural features to property were first reported in 1925 when Langmuir proposed to express the intermolecular interactions in liquid state using the surface energy.⁴ Accordingly, the surface energy of a hydrocarbon can be expressed in terms of its total molecular surface area because the surface energy per unit surface is assumed to be uniform.^{5,6} For polar molecules, Grigoras⁷ accounted for the nonuniform charge distribution due to heteroatoms by incorporating the sums of the positively and negatively charged atomic surface areas multiplied by the corresponding partial charges, as well as a hydrogen bonding term, as corrections to the total molecular surface area. An atomic charge scaling factor was introduced to correct the partial charges calculated by the extended Huckel theory. By using four especially derived composite parameters, the critical temperatures of 137 diverse organic compounds were correlated with R^2 of 0.946 and a standard error s of 16.4.

In a further advance, Jurs et al.⁸ introduced charged partial surface area (CPSA) descriptors, which combine solvent accessible surface areas with partial atomic charges; these descriptors allowed the accurate prediction of the critical temperatures of 147 diverse organic molecules with $R^2 = 0.978$ and $s = 11.9$ K using eight descriptors,⁹ in addition, 30 more compounds were successfully used as a test set, with $s = 13.23$. For the same set of 147 compounds, Hall et al.¹⁰ achieved $R^2 = 0.993$ and $s = 6.61$ K, employing 19 atom types of electrotopological state indices; an additional training set of 18 compounds gave $R^2 = 0.990$, with $s = 6.71$.

Using a model that combined five different connectivity indices, Seybold *et al.*¹¹ obtained an excellent correlation ($R^2 = 0.995$) for the critical temperatures of 74 normal and branched alkanes. In a set of 13 halogenated hydrocarbons, Seybold et al.¹² also successfully employed a Q_H descriptor that accounts for the charge on polar hydrogens in the regression model to achieve $R^2 = 0.964$ and 0.952 using a total of four and two descriptors, respectively. The approximation of charges on polar hydrogen was proposed by Di Paolo et al.¹³ A more detailed summary of the estimation methods for critical temperatures has been reported by Horvarth.¹⁴ Other literature has discussed various theories in QSPR for the estimation of liquid-state properties.^{15,16}

Although correlation equations involving many parameters may function as good tools for prediction, and even for revealing the crude microscopic aspects of structural features influencing the property, the *intrinsic* meaning of the descriptor space, as it relates to critical temperatures is not readily appreciated. In addition, when an excessive number of independent variables are involved in the correlation equation, overfitting and chance correlations may arise. Cramer¹⁷ attempted to account for the macromolecular properties in the liquid states using either two or up to five parameters obtained as principal components from six liquid-state properties (including critical temperatures) of 114 diverse compounds. Correlation of these two major components, “bulk” and “polar” cohesiveness, with 83 diverse compounds gave $R^2 = 0.902$.¹⁷ Politzer and his group¹⁸ used

two theoretical descriptors to correlate the critical temperatures of 65 diverse compounds resulting in a correlation coefficient of 0.914 ($R^2 = 0.835$) and a standard deviation of 59.7 K. One of these descriptors was the molecular surface area, and the other was $\nu\sigma_{\text{tot}}^2$, derived from the electrostatic potential calculated on molecular surfaces defined by Bader et al.¹⁹ The $\nu\sigma_{\text{tot}}^2$ measures the tendency of a molecule for electrostatic interactions. Recently, a statistical mechanics-based method, the bonded hard sphere theory, was used to calculate the critical temperatures of hydrocarbons.²⁰ The development of computer technology and its theoretical basis should ultimately allow QSPR to help build the foundations for calculating properties using molecular first principles.

Our research group has been pursuing QSPR models emphasizing revealing the intrinsic dimensionalities of properties while maintaining good predictive quality.^{21,22} We succeeded with two orthogonal descriptors in correlating the normal boiling points of 298 diverse organic compounds, achieving $R^2 = 0.9544$ and $s = 16.2$ K. One of the descriptors, $G_1^{1/3}$, resembled bulk cohesiveness whereas the other, the area-weighted surface charge of the hydrogen-bonding donor atom(s) in the molecule, described the ability of the molecule to hydrogen bond.²² This correlation equation also applies to the normal boiling points of a number of simple inorganic compounds including ammonia and water. It is extremely important to see if other liquid-state properties, such as critical temperatures, can also be described by the two descriptors, or, how far and why they account for the major variances of critical temperatures.

This paper is aimed at validating the two descriptors previously reported to represent the relevant inter- and intramolecular interactions in the liquid state. The relationship between critical temperatures and molecular structures is to be established using physically definite molecular descriptors in a limited parameter space.

METHODS, RESULTS, AND DISCUSSION

Critical temperatures of 165 important organic compounds drawn from the Design Institute for Physical Property Data (DIPPR)²³ database were collected as the complete set to develop the QSPR models. The same data set (147 compounds in their training set, plus 18 in their external predictions) has previously been used by Egolf et al.⁸ and Hall et al.⁹ This set is structurally very diverse and includes saturated and unsaturated hydrocarbons and halogenated compounds, and hydroxyl, cyano, amino, ester, ether, carbonyl, and carboxyl functionalities. The structures were drawn from scratch and pre-optimized by the molecular mechanics MMX method using the PCMODEL program.²⁴ The final geometry optimization of the molecules was performed on an IBM RISC/6000 model 320 using the semiempirical quantum-chemical AM1 method²⁵ within the MOPAC 6.0 program.²⁶ The MOPAC output files of individual compounds were loaded into the CODESSA for MS Windows program²⁷ along with the critical temperature data. The CODESSA program implements procedures that enable the calculation of a large selection of nonempirical descriptors as described in ref 22. Descriptors that are solely associated with a specific constituent, such as the number of C atoms, the minimum atomic state energy for a C atom,

Table 1. One-Parameter Correlations for the 76 Hydrocarbons

descriptor	R^2
3rd order Randic index, ${}^3\chi$	0.9224
gravitation index for all bonded pairs of atoms, G_1	0.8696
gravitation index for all pairs of atoms, G_p	0.8471
AM1 α -polarizability	0.8023
molecular weight, MW	0.7442
zero-order structural information content	0.7293
first-order Kier & Hall index, ${}^1\chi^v$	0.6258
Wiener index	0.3123

Table 2. One-Parameter Correlations for the 76 Hydrocarbons

descriptor	R^2
cube root of the gravitation index, $\sqrt[3]{G_1}$	0.9526
square root of the gravitation index, $\sqrt{G_1}$	0.9350
cube root gravitation index, for all pairs of atoms, $\sqrt[3]{G_p}$	0.9251
square root gravitation index for all pairs of atoms, $\sqrt{G_p}$	0.9143
cube root AM1 α -polarizability	0.8928
square root third-order Randic index, $\sqrt{{}^3\chi}$	0.8680
square root of the molecular weight, $\sqrt{\text{MW}}$	0.8193
cube root first-order Kier & Hall index	0.7302
cube root Wiener index	0.7016

and the average valence of a C atom, etc., were not included as they were considered to be of little physical significance. The correlation analysis to find the best QSPR model of a given size was carried out using two procedures described in ref 22. Both strategies yielded the same best correlations, which added confidence to the reliability of the methodology and the QSPR equations developed.

To avoid the initial complexities that arise from specific (mainly hydrogen-bonding) interactions between the liquid molecules, the QSPR treatment of the critical temperatures was commenced starting from the hydrocarbon subset of compounds. This procedure allows the elucidation of structural features attributable to "bulk" cohesiveness as described by Cramer.¹⁷ The hydrocarbon subset contained 76 C_3 – C_{19} alkanes, cycloalkanes, alkenes, alkylarenes, and alkynes. Several bulk-related descriptors performed well for the hydrocarbon set (Table 1). The best descriptor for the one-parameter correlations was the third-order Randic index. Other important descriptors included (i) the gravitational index over all bonded atoms, (ii) the gravitational index over all pairs of atoms, (iii) AM1 calculated α -polarizability of the molecule, (iv) the first-order Randic index,²⁸ (v) the molecular weight (MW), (vi) the structural information content,²⁹ and (vii) the first-order Kier and Hall index.¹⁵ A careful examination revealed that the critical temperature has some nonlinear dependence on those seven descriptors. Interestingly, when the square root or cube root of the descriptor was used, the correlations were significantly improved (Table 2). For instance, the Wiener index does not give a significant one-parameter correlation for the critical temperature of hydrocarbons unless the cube root of the descriptor is taken. The gravitational index simultaneously accounts for both the atomic masses (volumes) and for their distribution within the molecular space. Although the physical justification of the square and cube roots of descriptors is obscure, the observation indicates that the lower exponential order representation of the molecular bulk descriptors describes the related effective inter- and intramolecular interactions better. A similar trend has been reported for boiling points.^{22,30} The best one-parameter correlation for the hydrocarbon set was finalized as eq 1 (Figure 1). In

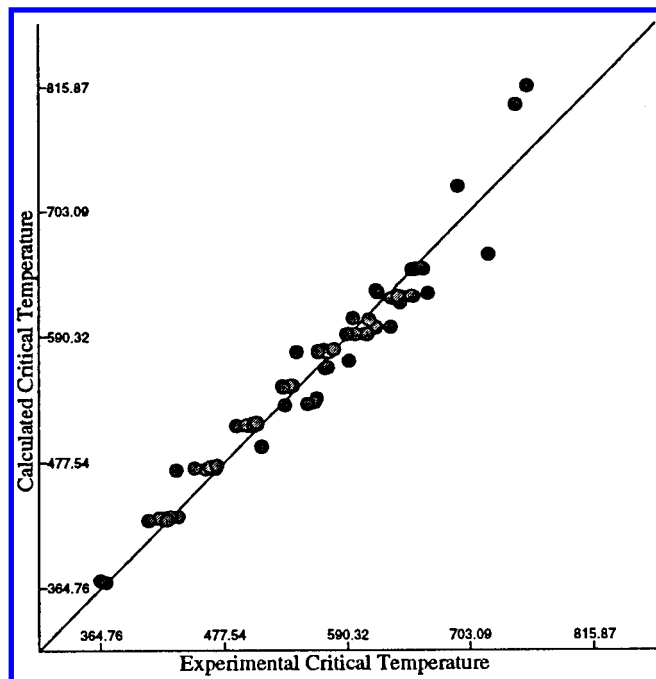


Figure 1. The calculated versus experimental critical temperatures of 76 hydrocarbons using the best one-parameter correlation equation (eq 1).

eq 2, which defines the gravitation index, m_i and m_j are the atomic masses of the bonded atoms and r_{ij} denotes the respective bond lengths:

$$T_c = (-100 \pm 17) + (79.7 \pm 2.1) \sqrt[3]{G_I} \quad (1)$$

$n = 76$, $R^2 = 0.9526$, $R_{cv}^2 = 0.9472$, $F = 1506$, $s = 18.9$, averaged absolute error = 13.9

$$G_I = \sum_{i,j}^{\text{all bonded pairs}} \frac{m_i m_j}{r_{ij}^2} \quad (2)$$

For the complete data set of 165 compounds, the dominant influence of the gravitation index was moderated because of the involvement of specific interactions (mainly hydrogen-bonding self-association) for the majority of the compounds. The best one-parameter correlation with the cube root of G_I yielded a R^2 of only 0.7191, which was substantially poorer than those of the corresponding correlations for the hydrocarbon sets. Accordingly, a significant improvement of the correlations for the whole data set was observed if the hydrogen-bonding related descriptors were added to the QSPR equation. The best two-parameter equation (eq 4) involved the cube root of G_I and the area-weighted surface charge of the hydrogen bonding donor atom(s) in the molecule, where q_D is the partial charge on hydrogen bonding donor (H) atom(s), S_D denotes the surface area for this atom, and S_{tot} is the total molecular surface area, calculated from the van der Waals' radii of the atoms (overlapping spheres). The summation in eq 5 was performed over the number of simultaneously possible hydrogen bonding donor and acceptor pairs per molecule. Also, the hydrogen atoms at the α -position to carbonyl and cyano groups were accounted as possible hydrogen bonding donor with much smaller strength measured by significantly smaller partial charges on them:

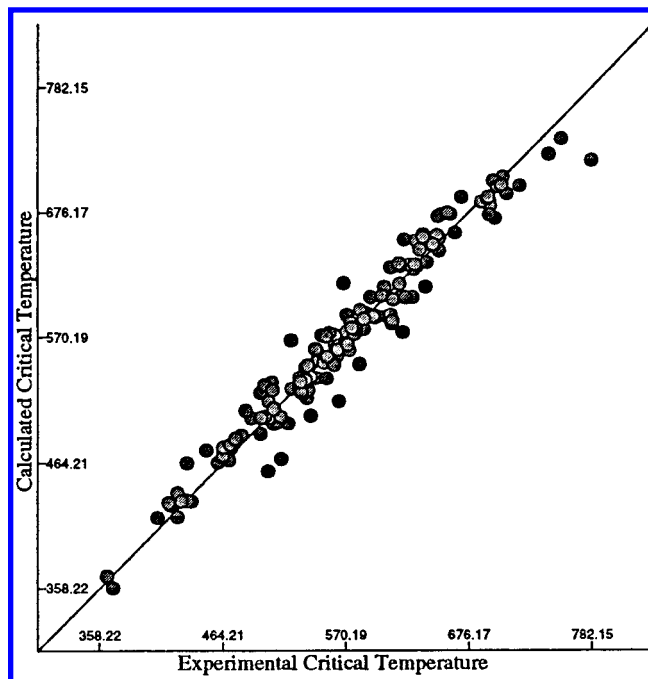


Figure 2. The calculated versus experimental critical temperatures of 165 diverse compounds using the best three-parameter correlation equation (eq 6).

$$T_c = (-38.6 \pm 29.5) + (74.4 \pm 3.6) \sqrt[3]{G_I} \quad (3)$$

$n = 165$, $R^2 = 0.7191$, $R_{cv}^2 = 0.7167$, $F = 428$, $s = 41.8$

$$T_c = -(109.1 \pm 20.9) + (80.6 \pm 2.5) \sqrt[3]{G_I} + (21388 \pm 1563) HDSA(2) \quad (4)$$

$n = 165$, $R^2 = 0.8680$, $F = 546$, $s = 28.7$

$$HDSA(2) = \sum \frac{q_D \sqrt{S_D}}{S_{tot}} \quad (5)$$

By adding the third descriptor, the topographic electronic index (T_E),³¹ the correlation was improved to $R^2 = 0.9550$ (eq 6, Figure 2). In eq 7, q_i and q_j are the Zefirov partial charges³² of the atoms, and r_{ij} is the respective bond length. Equation 7 itself does not explain the physical mechanism by which it is connected with the critical temperatures in the liquid state. However, further investigation showed that the topographic electronic index is correlated to the AM1 calculated polarizability of the molecule (the intercorrelation $R^2 = 0.6633$). The negative sign in front of T_E peculiarly indicates that increasing polarizability of a compound lowers the critical temperature. In fact, the highest absolute T_E values came from molecules containing hydrogen-bonding donors, and the lowest ones are recognized from molecules with nitrile group(s) alkenes and aromatic compounds. The dispersion and hydrogen-bonding ability of the latter group of compounds are not well accounted for by the gravitation index and the surface area weighed hydrogen-bonding donor charges. In eq 4, the hydrogen-bonding self-association is solely accounted for by the donors, which indeed play a key role. However, the differences in surface area among the acceptor sites, such as nitrile groups versus all other acceptors, do affect the hydrogen-bonding ability. The

Table 3. Verification of Statistical Validity

set	R^2	R_{cv}^2	S
I	0.9641	0.9572	15.6
II	0.9570	0.9493	18.9
III	0.9606	0.9550	13.4

negative T_E term may arise mainly because of the inclusion of the nitriles and thus represents merely a correction of the first two descriptors in eq 6. Notably, the replacement of T_E by any one of the CPSA descriptors for hydrogen-bonding acceptors reduces the quality of the correlation shown in eq 8, in which the hydrogen-bonding acceptor changed area (HACA) was used:

$$T_c = -(204 \pm 14) + (100.4 \pm 1.8)\sqrt[3]{G_1} + (27400 \pm 980)HDSA(2) - (69.3 \pm 3.9)T_E \quad (6)$$

$$n = 165, R^2 = 0.9550, R_{cv}^2 = 0.9547, F = 1166, s = 16.8, \text{ averaged absolute error} = 14.5$$

$$T_E = \sum_{i < j}^{\text{all pairs}} \frac{|q_i - q_j|}{r_{ij}^2} \quad (7)$$

$$T_c = -(103 \pm 23) + (79.7 \pm 2.8)\sqrt[3]{G_1} + (19500 \pm 1700)HDSA(2) + (27500 \pm 7800)HACA \quad (8)$$

$$n = 165, R^2 = 0.8829, F = 415, s = 27.1$$

$$HACA(2) = \sum \frac{q_A \sqrt{S_A}}{S_{tot}} \quad (9)$$

Thus, the three descriptors in eq 6 seem to account for (i) the dispersion and cavity-formation effects in liquids, (ii) the ability of compounds to hydrogen bond, and (iii) inadequacies in the previous two descriptors to model the critical temperatures of our particular set of compounds by correcting for the influence of hydrogen-bonding acceptors and branching effects in different isomers. It is interesting to note that the corresponding correction for the boiling points involved different descriptors.²² This result supports an early report that suggested that the descriptors that account individually for shape-dependent variations in the liquid-state properties are property specific.²³

To further demonstrate the absence of chance correlations, we divided the whole set into three subsets (by using nos 1, 4, 7....; 2, 5, 8....; and 3, 6, 9....) and predicted each subset by using the other two subsets as the training set. The results are shown in Table 3.

The three descriptors of eq 6 are almost orthogonal because the intercorrelation coefficients between any two of them is <0.2 . Equation 6 demonstrates that three descriptors, of which two closely resemble the widely recognized concept of dispersion and hydrogen-bonding self-association in the liquid state, describe most of the variance of the critical temperatures for a wide variety of organic substances. Predictions of critical temperatures for new compounds employing theoretically justified descriptors give greater confidence than using those that have less apparent meaning in the context of the property.

Besides physically justifying the descriptors, the statistical quality of the models developed here is competitive. For the hydrocarbon case, the precision of the one-parameter model is substantially better than the two-parameter model developed from 20 hydrocarbons by Murray et al.¹⁸ The three-parameter model for diverse compounds shows precision similar to the four-parameter model developed by Grigoras⁷; in addition, our model involves none of the empirical scaling of partial charges that Grigoras' model does. In comparison with the eight-parameter model developed for the subset (147 compounds) by Egolf and Jurs et al.,⁹ we used less than half the number of descriptors to represent the two well-recognized liquid-state intermolecular interactions (i.e., size-dependent dispersion and cavity formation as well as hydrogen-bonding self-association) at small cost in precision. Table 4 compares the experimental and calculated critical temperatures using eqs 1 and 6 for hydrocarbons and diverse compounds, respectively. No formulas for calculating the descriptors are provided in refs 7 and 9 and it is thus impossible to find out how the three leading descriptors in each of these references would perform in the critical temperature correlation of the 165 compounds.

As expected, by adding more descriptors in eq 6, the statistical quality can be further enhanced. As shown in eq 10, inclusion of the 3rd-order Randic index improves the R^2 to 0.9604, ($R = 0.9800$), and reduces the standard error s to 15.8 K (absolute error = 11.3 K). Further addition of the 1st-order information content index (1inf) (see eq 11) refines R^2 to 0.9651 ($R = 0.9824$), with the standard error s reduced to 14.9 K (absolute error = 10.5 K). More descriptors could be added to improve the R^2 value in this way. However, such improvement of the statistics is of doubtful physical significance because the 3rd-order Randic index ($^3\chi$) in eq 10 already partially compensates for the imperfection of the gravitation index by accounting for branching in the molecules. The same arguments are true for the involvement of the 1st-order information content index. Even though a better statistical model is obtained adding more descriptors, we believe that it would be less meaningful in revealing the intrinsic relationship between the property of interest and the chemical structure.

$$T_c = -(126 \pm 25) + (87.4 \pm 3.9)\sqrt[3]{G_1} + (26900 \pm 920)HDSA(2) - (61.4 \pm 4.1)T_E + (13.2 \pm 3.5)^3\chi \quad (10)$$

$$n = 165, R^2 = 0.9604, R_{cv}^2 = 0.9578, F = 995.256, s = 15.8, \text{ absolute error} = 11.3$$

$$T_c = -(147 \pm 24) + (92.5 \pm 3.8)\sqrt[3]{G_1} + (27600 \pm 880)HDSA(2)$$

$$- (57.8 \pm 4.0)T_E + (15.9 \pm 3.4)^3\chi - (0.85 \pm 0.18)^1inf \quad (11)$$

$$n = 165, R^2 = 0.9651, R_{cv}^2 = 0.9624, F = 902.0, s = 14.9, \text{ absolute error} = 10.5$$

The dispersion and cavity formation energies are often related to the size (e.g., the surface area) of the solute molecule. For instance, Cramer and Truhlar³⁵ employed in their AM1-SMx solvation theories the following formula-

Table 4. Comparison of the Experimental and Predicted Critical Temperatures (K) for 165 Diverse Compounds^a and 76 Hydrocarbons^b

no.	compound	exptl	calcd ^a	abs error	calcd ^b	abs error	no.	compound	exptl	calcd ^a	abs error	calcd ^b	abs error
1	3-chloropropene	514.2	467.2	47.0			84	2,3-dimethylpentane	537.4	532.8	4.6	545.0	7.6
2	propylene	364.8	368.3	3.5	371.1	6.4	85	<i>n</i> -heptane	540.3	536.0	4.3	545.7	5.4
3	1,2-dichloropropene	545.1	535.1	10.0			86	2-methylhexane	530.4	535.6	5.2	545.4	15.0
4	ethylformate	508.4	496.4	12.0			87	3-methylhexane	535.3	533.5	1.8	545.4	10.1
5	methyl acetate	506.8	497.6	9.2			88	ethylbenzene	617.2	629.9	12.8	597.6	19.6
6	propionic acid	604.0	612.6	8.6			89	<i>m</i> -xylene	617.1	632.2	15.2	598.4	18.7
7	<i>n</i> -propyl chloride	503.2	456.7	46.5			90	<i>p</i> -xylene	616.3	632.4	16.1	598.4	17.9
8	propane	369.8	358.2	11.6	369.5	0.4	91	2,6-xyleneol	701.1	697.4	3.6		
9	methyl ethyl ether	437.8	431.2	6.6			92	<i>N,N</i> -dimethylaniline	687.2	684.9	2.2		
10	<i>n</i> -propanol	536.7	518.3	18.4			93	<i>cis</i> -1,2-dimethylcyclohexane	606.2	587.6	18.6	592.0	14.0
11	methylal	480.6	487.0	6.4			94	<i>trans</i> -1,2-dimethylcyclohexane	596.2	588.1	8.1	591.9	4.3
12	isopropylamine	471.9	476.1	4.3			95	<i>cis</i> -1,3-dimethylcyclohexane	591.2	589.2	1.9	592.2	1.0
13	<i>n</i> -propylamine	497.0	488.3	8.7			96	<i>trans</i> -1,3-dimethylcyclohexane	598.0	587.8	10.2	591.8	6.2
14	trimethylamine	433.3	432.7	0.6			97	<i>trans</i> -1,4-dimethylcyclohexane	590.2	587.8	2.4	591.9	1.7
15	1,3-butadiene	425.4	438.6	13.2	427.8	2.4	98	ethylcyclohexane	609.2	589.0	20.1	591.6	17.6
16	butyronitrile	582.3	546.9	35.4			99	isobutyl isobutyrate	602.0	605.3	3.3		
17	1-butene	419.6	428.4	8.8	426.8	7.2	100	2,3-dimethylhexane	563.4	562.0	1.4	576.8	13.4
18	<i>cis</i> -2-butene	435.6	431.2	4.3	428.4	7.2	101	2-methyl-3-ethylpentane	567.0	560.0	7.0	576.4	9.4
19	<i>trans</i> -2-butene	428.6	431.3	2.6	428.2	0.4	102	<i>n</i> -octane	568.8	567.7	1.1	577.9	9.1
20	isobutene	417.9	430.0	12.1	427.3	9.4	103	2,2,3-trimethylpentane	563.5	559.0	4.5	575.5	12.0
21	methyl ethyl ketone	535.5	544.1	8.6			104	2,2,4-trimethylpentane	544.0	559.4	15.4	576.0	32.0
22	tetrahydrofuran	540.2	503.2	36.9			105	2,3,3-trimethylpentane	573.5	558.6	14.9	575.7	2.2
23	<i>n</i> -butyric acid	628.0	603.7	24.3			106	2-ethyl-1-hexanol	640.3	633.7	6.6		
24	ethyl acetate	523.3	526.0	2.7			107	quinoline	782.2	719.8	62.3		
25	isobutyric acid	609.2	629.2	20.1			108	cumene	631.2	651.8	20.7	624.4	6.8
26	methyl propionate	530.6	526.8	3.8			109	<i>o</i> -ethyltoluene	651.2	653.8	2.7	625.5	25.7
27	<i>n</i> -propyl formate	538.0	525.1	12.9			110	<i>p</i> -ethyltoluene	640.2	655.1	14.9	625.6	14.6
28	<i>n</i> -butane	425.2	418.2	7.0	425.4	0.2	111	mesitylene	637.4	657.3	19.9	626.2	11.2
29	isobutane	408.1	417.6	9.5	424.9	16.8	112	<i>n</i> -propylbenzene	638.4	653.6	15.2	625.1	13.3
30	butanol	562.9	554.1	8.8			113	1,2,3-trimethylbenzene	664.5	658.3	6.2	628.4	36.1
31	<i>sec</i> -butyl alcohol	536.0	537.9	1.9			114	1,2,4-trimethylbenzene	649.1	656.8	7.7	626.1	23.0
32	<i>tert</i> -butyl alcohol	506.2	531.6	25.4			115	<i>n</i> -propylcyclohexane	639.2	612.6	26.5	620.0	19.2
33	diethyl ether	466.7	469.8	3.1			116	3,3-diethylpentane	610.1	581.3	28.7	604.9	5.2
34	isobutanol	547.7	550.3	2.6			117	<i>n</i> -nonane	595.7	586.6	9.0	606.0	10.4
35	<i>n</i> -butylamine	531.9	522.9	9.0			118	2,2,3,3-tetramethylpentane	610.9	583.8	27.1	604.6	6.2
36	diethylamine	496.6	522.8	26.2			119	1,2,3,4-tetrahydronaphthalene	720.2	698.5	21.6	663.4	56.8
37	pyridine	620.0	573.9	46.1			120	<i>n</i> -butylbenzene	660.6	674.5	13.9	650.5	10.1
38	cyclopentane	511.8	498.3	13.5	491.1	20.7	121	<i>p</i> -cymene	653.2	674.4	21.2	650.3	2.9
39	1-pentene	464.8	475.4	10.6	472.6	7.8	122	isobutylbenzene	650.2	672.4	22.3	649.9	0.3
40	diethyl ketone	561.0	562.9	1.9			123	<i>n</i> -decane	618.5	604.7	13.7	629.3	10.8
41	2-pentanone	561.1	572.0	10.9			124	<i>n</i> -tetradecane	692.4	687.5	4.9	725.0	32.6
42	ethyl propionate	546.0	550.4	4.4			125	<i>n</i> -octadecane	745.3	725.4	19.9	799.0	53.7
43	isobutyl formate	551.4	548.2	3.1			126	<i>n</i> -nonadecane	755.9	738.5	17.4	815.9	60.0
44	<i>n</i> -propyl acetate	549.4	571.4	22.0			127	<i>sec</i> -butyl chloride	520.6	496.8	23.8		
45	valeric acid	651.0	643.5	7.5			128	methyl isopropyl ether	464.5	469.6	5.1		
46	isopentane	460.4	464.2	3.8	471.0	10.6	129	<i>sec</i> -butylamine	514.3	502.3	12.0		
47	neopentane	433.8	463.5	29.7	470.0	36.2	130	<i>tert</i> -butylamine	483.9	507.7	23.8		
48	<i>n</i> -pentane	469.7	466.0	3.6	471.5	1.8	131	2-methyl-1-butene	465.0	476.9	11.9	473.1	8.1
49	2-methyl-2-butanol	545.2	559.1	13.9			132	2-methyl-2-butene	471.0	479.4	8.4	474.3	3.3
50	3-methyl-1-butanol	579.5	574.2	5.3			133	3-methyl-1-butene	450.4	474.6	24.2	471.8	21.4
51	1-pentanol	586.2	576.6	9.5			134	methyl isopropyl ketone	553.0	570.3	17.3		
52	bromobenzene	670.2	688.7	18.5			135	methyl <i>n</i> -butyrate	554.5	553.2	1.3		
53	chlorobenzene	632.4	630.5	1.9			136	piperidine	594.1	587.2	6.8		
54	benzene	562.2	571.7	9.5	534.7	27.5	137	ethyl propyl ether	500.2	502.7	2.5		
55	phenol	694.3	685.1	9.1			138	methyl <i>tert</i> -butyl ether	497.1	501.5	4.4		
56	aniline	699.0	671.0	28.0			139	3-methylpyridine	645.0	648.0	3.0		
57	2-methylpyridine	621.0	652.7	31.7			140	1,5-hexadiene	507.0	525.2	18.2	512.7	5.7
58	cyclohexene	560.4	546.5	13.9	531.0	29.4	141	hexanenitrile	622.1	603.6	18.5		
59	cyclohexanone	629.2	627.3	1.8			142	3-hexanone	582.8	593.1	10.3		
60	cyclohexane	553.5	535.0	18.5	529.4	24.8	143	ethyl isobutyrate	553.2	570.8	17.6		
61	1-hexene	504.0	515.4	11.4	512.0	8.0	144	<i>n</i> -propyl propionate	578.0	572.3	5.7		
62	methylcyclopentane	532.8	533.3	0.5	528.1	4.7	145	1-hexanol	611.4	602.0	9.3		
63	cyclohexanol	625.2	631.3	6.1			146	2-hexanol	586.2	585.4	0.8		
64	2-hexanone	587.1	590.8	3.8			147	4-methyl-2-pentanol	574.4	582.3	7.9		
65	methyl isobutyl ketone	571.4	588.7	17.3			148	acetone	508.2	509.9	1.7		
66	ethyl <i>n</i> -butyrate	571.0	574.6	3.6			149	2-propanol	508.3	509.0	0.7		
67	isobutyl acetate	561.0	571.5	10.5			150	ethyl vinyl ether	475.2	484.6	9.4		
68	2,2-dimethylbutane	488.8	501.2	12.4	509.6	20.8	151	isovaleric acid	634.0	644.1	10.1		
69	2,3-dimethylbutane	500.0	501.6	1.6	510.1	10.1	152	<i>n</i> -butyl acetate	579.2	576.8	2.4		
70	<i>n</i> -hexane	507.4	503.2	4.3	510.8	3.4	153	<i>p</i> -cresol	704.7	699.1	5.5		
71	2-methylpentane	497.5	502.0	4.5	510.1	12.6	154	<i>p</i> -toluidine	693.2	689.0	4.2		
72	3-methylpentane	504.4	501.7	2.7	510.7	6.3	155	1-heptene	537.3	545.6	8.3	545.8	8.5
73	diisopropyl ether	500.1	529.2	29.1			156	methylcyclohexane	572.2	563.2	9.0	562.3	9.9
74	di- <i>n</i> -propyl ether	530.6	531.8	1.2			157	2,2,3-trimethylbutane	531.2	531.9	0.8	544.0	12.8
75	diisopropylamine	523.1	567.9	44.8			158	<i>o</i> -xylene	630.4	631.6	1.3	598.2	32.2
76	di- <i>n</i> -propylamine	555.8	573.3	17.5			159	1-octene	578.2	577.9	0.3	578.6	0.4
77	benzaldehyde	695.0	680.8	14.2			160	<i>m</i> -methyltoluene	637.2	655.0	17.8	625.6	11.6
78	toluene	591.8	603.8	12.0	568.0	23.8	161	<i>p</i> -diethylbenzene	658.0	675.7	17.7	651.0	7.0
79	<i>m</i> -cresol	705.9	706.1	0.3			162	1-decene	617.1	615.1	2.0	630.7	13.6
80	<i>o</i> -cresol	697.6	702.8	5.3			163	propionitrile	564.4	515.7	48.7		
81	<i>m</i> -toluidine	709.2	691.4	17.7			164	4-methylpyridine	646.2	648.4	2.3		
82	<i>o</i> -toluidine	694.2	673.9	20.3			165	<i>n</i> -pentyl formate	576.0	577.7	1.7		
83	ethylcyclopentane	569.5	563.0	6.5	561.4	8.1							

Table 5. The Gravitation Index and the HDCA(2) in One- and Two-Parameter Correlations with Critical Temperatures, Boiling Points, and Ostwald Solubility Coefficients

property	descriptor	R^2
critical temperatures ^a	$\sqrt[3]{G_1}$	0.7191
	$\sqrt[3]{G_1} + \text{HDCA}(2)$	0.8680
normal Boiling Points ^b	$\sqrt[3]{G_1}$	0.7718
	$\sqrt[3]{G_1} + \text{HDCA}(2)$	0.9544
Ostwald solubility coefficients ^c	$\sqrt[3]{G_1}$	0.8717
	$\sqrt[3]{G_1} + \text{HDCA}(2)$	0.9012

^a 165 diverse organic compounds from this study. ^b 298 diverse organic compounds (ref 22). ^c Diverse set of 394 gases and vapors according to availability (refs 3 and 4).

tion of the cavity formation free energy: $\Delta G(\text{cav}) = \sigma^* S(M)$, where sigma is the surface tension of the solvent and $S(M)$ is the solvent-accessible surface area of the solute. Tomasi and others³⁶ have developed a scheme that reduces the spatial representation of the dispersion interaction into a surface representation of this interaction, thus effectively connecting the size of the dispersion effect with the solvent-accessible surface area of the solute.

The definition of the gravitation index reflects the effective mass distribution in the molecule and thus directly represents the "effective" size of the molecule. It has been established by our treatment of (i) boiling points points,²² (ii) gas solubilities in aqueous solutions,^{21d} and (iii) critical temperatures (present work) that the gravitation index (or an exponent) is by far the most important (leading) term in the QSPR description of the relevant solute-solvent interactions, particularly in the case of non-hydrogen bonding solvents. Being directly related to the molecular size, it is thus theoretically justified to ascribe the effect accounted for by the gravitation index to dispersion and cavity formation. (Electrostatic solute-solvent interactions and hydrogen-bonding effects are not explicitly related to the size of the solute molecule.)

As reported earlier,²² the two leading descriptors, the gravitation index and the surface area weighted hydrogen-bonding donor charges, account for 95% of the variance of normal boiling points of 298 diverse organic compounds (Table 5). These two descriptors were also used to correlate the Ostwald solubility coefficients of diverse gases and vapors in hexadecane, resulting in $R^2 = 0.9012$ (Table 4). Thus, in all three cases (critical temperatures, boiling points, and Ostwald solubility coefficients), these two descriptors consistently represent the major intrinsic dimensions of dispersion and hydrogen-bonding interactions. The gravitation index is advantageous because it encodes not only the information of mass, but mass distribution in the molecular space. It appropriately illustrates that cyclic compounds with lower molecular weights than their chain derivatives have higher critical temperatures because their gravitation indices are higher. For example, cyclohexane with the molecular weight of 84 g/mol has a critical temperature of 553.5 K, whereas *n*-hexane with a molecular weight of 86 g/mol has a critical temperature 46 K degrees lower. The same trend follows for boiling points.

CONCLUSION

One- and three-parameter QSPR models were developed for the critical temperatures of sets of hydrocarbons and of

diverse molecules. The one-parameter model allows the prediction of critical temperatures for hydrocarbons with an average error of 13.9 K, whereas the three-parameter predictions of critical temperatures for diverse molecules have an average error of 14.8 K. The models reconfirmed that the molecular size-dependent bulk effects (dispersion and cavity formation) in the liquid state can be represented by functions of the gravitation index, whereas the hydrogen-bonding self-association interactions can be represented by the area weighted surface charge of the hydrogen-bonding donor atoms. However, this structural features of donor hydrogen alone does not account completely for differences among various hydrogen-bonding acceptors. This inadequacy is more serious for the critical temperature than for the boiling point. The supplementary descriptors needed to account for the differences of hydrogen-bonding acceptors and branching effects in different isomers differ for the two properties.

REFERENCES AND NOTES

- (1) Meissner, H. P. Critical Constants from para-Chor and Molar Refraction. *Chem. Eng. Progr.* **1949**, *45*, 149–153.
- (2) Joback, K. G.; Reid, R. C. Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.* **1987**, *57*, 233–243.
- (3) Constantinou, L.; Gani, R. New Group Contribution Method for Estimating Properties of Pure Compounds. *Thermodynamics* **1994**, *40*, 1697–1710.
- (4) Langmuir, I. In *Colloid Symp. Monograph*; Holmes, H. N., Ed.; The Chemical Catalog Company: Baltimore, 1925; p 48.
- (5) Hermann, R. B. Theory of Hydrophobic Bonding. II The Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area. *J. Phys. Chem.* **1972**, *76*, 2754–2759.
- (6) Hermann, R. B. Theory of Hydrophobic Bonding. III A Method for the Calculation of the Hydrophobic Interaction Based on Liquid-State Perturbation Theory and a Simple Liquid Model. *J. Phys. Chem.* **1975**, *79*, 163–169.
- (7) Grigoros, S. A Structural Approach to Calculated Physical Properties of Pure Organic Substances: The Critical Temperature, Critical Volume and Related Properties. *J. Comput. Chem.* **1990**, *11*, 493.
- (8) 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.
- (9) 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.
- (10) Hall, L. H.; Story, C. T. Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electropotential State Indices Using Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004–1014.
- (11) 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.
- (12) Seybold, P. G.; May, M. A.; Gargas, M. L. On the Physical Properties of Halogenated Hydrocarbons. *Acta Pharm. Jugosl.* **1986**, *36*, 253–265.
- (13) Paolo, D. T.; Kier, L. B. Molecular Connectivity and Structure-Activity Relationship of General Anesthetics. *Mol. Pharmacol.* **1977**, *31*, 31–37.
- (14) Horvath, A. L. *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds*; Horvath, A. L., Ed.; Elsevier: Amsterdam, 1992; Chapter 2, p 131.
- (15) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; John Wiley & Sons: New York, 1986.
- (16) (a) Balaban, A. T. *Chemical Applications of Graph Theory*; Academic: New York, 1976; (b) Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H. *Handbook of Chemical Property Estimation Methods*; McGraw-Hill: New York, 1982; (c) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC: Boca Raton, FL, 1992.
- (17) Cramer, R. D., III BC(DEF) Parameters. I. The Intrinsic Dimensionality of Intermolecular Interaction in the Liquid State. *J. Am. Chem. Soc.* **1980**, *102*, 1837–1859.
- (18) Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, E. M.; Politzer, P. Relationships of Critical Constants and Boiling Points to Computed Molecular Surface Properties. *J. Phys. Chem.* **1993**, *97*, 9369–9373.

- (19) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. Properties of Atoms in Molecules: Atomic Volumes. *J. Am. Chem. Soc.* **1987**, *109*, 7968–7979.
- (20) Archer, A. L.; Amos, M. D.; Jackson, G.; McLure, I. A. The Theoretical Prediction of the Critical Points of Alkanes, Perfluoroalkanes, and Their Mixtures Using Bonded Hard-Sphere (BHS) Theory. *Int. J. Thermophys.* **1996**, *17*, 201–210.
- (21) (a) Murugan, R.; Grendze, M. P.; Toomey, J. E. Jr.; Katritzky, A. R.; Karelson, M.; Lobanov, V. S.; Rachwal, P. Predicting Physical Properties from Molecular Structure. *CHEMTECH* **1994**, *24*, 17–19; (b) 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; (c) Katritzky, A. R.; Rachwal, P.; Law, K. W.; Karelson, M.; Lobanov, V. Prediction of Polymer Glass Transition Temperatures Using a General Quantitative Structure–Property Relationship Treatment. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 879–884; (d) Katritzky, A. R.; Mu, L.; Karelson, M. A QSPR Study of the Solubility of Gases and Vapors in Water. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1162–1168; (e) Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure–Property Relationship Approach. *Langmuir* **1996**, *12*, 1462–1470; (f) Katritzky, A. R.; Lobanov, V. S.; Karelson, M.; Murugan, R.; Grendze, M. J.; Toomey, J. E., Jr. Comprehensive Descriptors for Structural and Statistical Analysis. 1 Correlations Between Structure and Physical Properties of Substituted Pyridines. *Rev. Roum. Chim.* **1996**, *41*, 851–867; (g) Katritzky, A. R.; Mu, L.; Karelson, M. QSPR Treatment of the Unified Non-Specific Solvent Polarity Scale. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 756–761; (h) Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S., in preparation.
- (22) 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.
- (23) Daubert, T. E.; Danner, R. P. *Physical and Thermodynamic Properties of Pure Chemicals: Data Compilation*; Design Institute for Physical Property Data (DIPPR). Hemisphere: New York, 1989; Vol. 1–4.
- (24) *PCMODEL*, 5th Ed.; Serena Software: Bloomington, IN, 1992.
- (25) 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.
- (26) Stewart, J. J. P. *MOPAC Program Package 6.0*, QCPE, No. 455, 1990.
- (27) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA*, User's Manual, University of Florida & Tartu University, 1994.
- (28) Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
- (29) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indices in Organic Chemistry. *Russ. Chem. Rev.* **1988**, *57*, 191–203.
- (30) (a) Walker, J. The Boiling Points of Homologous Compounds. Part I. Simple and Mixed Ethers. *J. Chem. Soc.* **1894**, *65*, 193–202; (b) Kinney, C. R. The Mechanism of Aromatic Bromination. *J. Am. Chem. Soc.* **1938**, *60*, 3032–3039; (c) Kamlet, M. J.; Doherty, R. M.; Taft, R. W.; Abraham, M. H.; Koros, W. J. Solubility Properties in Polymers and Biological Media. 3. Predictional Methods for Critical Temperatures, Boiling Points, and Solubility Properties (R_G Values) Based on Molecular Size, Polarizability, and Dipolarity. *J. Am. Chem. Soc.* **1984**, *106*, 1205–1212; (d) 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.
- (31) Osmialowski, K.; Halkiewicz, J.; Kaliszan, R. Quantitative Chemical Parameters in Correlation Analysis of Gas–Liquid Chromatographic Retention Indices of Amines. *J. Chromatogr.* **1986**, *361*, 63–69.
- (32) 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 (Engl. Transl.)* **1987**, *296*, 883–887.
- (33) Randic, M.; Seybold, P. G. Molecular Shape as a Critical Factor in Structure–Property–Activity Studies. *SAR and QSAR in Environmental Research* **1993**, *1*, 77–85.
- (34) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors that Influence the Solubility of Gases and Vapours in Water at 298K, and a New Method for Its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.
- (35) Cramer, C. J.; Truhlar, D. G. General Parametrized SCF Model for Free Energies of Solvation in Aqueous Solution. *J. Am. Chem. Soc.* **1991**, *113*, 8305–8311.
- (36) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev. (Washington, D.C.)* **1994**, *94*, 2027–2094.

CI970071Q