Annual Review

QUANTITATIVE STRUCTURE–PROPERTY RELATIONSHIPS FOR PREDICTION OF BOILING POINT, VAPOR PRESSURE, AND MELTING POINT

John C. Dearden*

School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, United Kingdom

Abstract—Boiling point, vapor pressure, and melting point are important physicochemical properties in the modeling of the distribution and fate of chemicals in the environment. However, such data often are not available, and therefore must be estimated. Over the years, many attempts have been made to calculate boiling points, vapor pressures, and melting points by using quantitative structure–property relationships, and this review examines and discusses the work published in this area, and concentrates particularly on recent studies. A number of software programs are commercially available for the calculation of boiling point, vapor pressure, and melting point, and these have been tested for their predictive ability with a test set of 100 organic chemicals.

Keywords—Quantitative structure–property relationships   Prediction   Boiling point   Vapor pressure   Melting point

INTRODUCTION

The environmental impact of a chemical is dictated by its physicochemical properties, which include among others, boiling point, vapor pressure, and melting point. The first two are, of course, related, because boiling point is the temperature at which vapor pressure equals atmospheric pressure; in fact, boiling point is used in some quantitative structure–property relationship (QSPR) models to predict vapor pressure. These properties are important because the vapor pressure of a chemical affects its residence time in soil and in water, and thus is a significant factor in its distribution and transport in the environment, being used in fugacity calculations [1]. Vapor pressure also has been used as a descriptor in quantitative structure–activity relationships (QSARs) to predict toxicity. Melting point, too, has been used to predict vapor pressure, and is valuable also in that it indicates the physical state of the chemical at ambient temperatures, which will dictate how the chemical is handled and treated. In addition, melting point can be used in the prediction of aqueous solubility [2].

Nevertheless, even basic physicochemical properties often are not available for many chemicals that are released into the environment. Furthermore, new chemical entities often are submitted to the regulatory authorities without even such simple-to-determine properties as melting point being provided. Thus, it is valuable to be able to predict these properties by using QSPR techniques, which have the additional advantage of being quick to use.

A QSPR is a mathematical description of a property in terms of other properties (descriptors) that are of three broad classes—hydrophobic, electronic, and steric. Since the advent of modern QSAR investigations in 1962 [3], many attempts to model physicochemical properties have been published. It is worth mentioning that many QSPRs predate this period, with the earliest QSPRs for melting point and boiling point being published as long ago as 1884 by Mills [4]. His QSPRs were for homologous series, and took the form $t = \beta x/(1 + \gamma x)$,

where t is the temperature (°C), x is the number of methylene units in the hydrocarbon chain, and β and γ are constants for a given chemical class. His predictions were remarkably accurate, with errors generally being $< 1°$ for both melting and boiling points.

This paper reviews the QSPR prediction of boiling point, vapor pressure, and melting point, and also reports the testing of a number of commercial software programs that predict these properties. Earlier reviews dealt with each of these areas, so we shall concentrate here on recent studies.

BOILING POINT

Normal boiling point is defined as the temperature at which a substance has a vapor pressure of 760 mm Hg. Boiling point is a function of a number of molecular properties that control the ability of a molecule to escape from the surface of a liquid into the vapor phase. These properties are molecular size (which controls dispersion interactions within the liquid phase, and also dictates the energy required to create a cavity in the liquid), and polar and hydrogen bonding forces. Entropic factors, such as molecular flexibility and orientation, also affect boiling point, although the entropy of vaporization is reasonably constant for small, rigid, non–hydrogen-bonded molecules (Trouton's rule). Zhao et al. [5] developed a QSPR for the entropy of boiling, based on molecular symmetry and flexibility, and group contributions.

Most studies concerned with boiling point prediction have used normal boiling points, that is, those determined at 760 mm Hg pressure. Although experimental boiling points would be expected to be highly accurate, it should be noted that the presence of impurities and thermal decomposition (especially for compounds with high boiling points) can cause errors. Moreover, as Rücker and Rücker [6] pointed out, literature values are not necessarily accurate; they found wide discrepancies even for alkanes, with the boiling point of cyclooctane reported from 120.3 to 156°C, and that for methylcycloheptane from 113 to 136°C.

Horvath [7], Nendza [8], and Lyman [9] summarized early

* j.c.dearden@livjm.ac.uk.

work on boiling point prediction. Nendza [8] stated that the recommended method for many monofunctional compounds is that based on the work of Meissner [10] and Rechsteiner [11], which relates boiling point (T_b; in degrees kelvin [K]) to molar refractivity (MR), parachor (*P*), and a class-dependent variable *B*

$$T_b = [637(MR)^{1.47} + B]/P \qquad (1)$$

Lyman [9] discussed seven recommended methods for boiling point estimation, namely those of Banks [12], a simple correlation with the square root of molecular weight (MW); Burnop [13], a correlation with the square root of molecular weight and structural factors; Cramer [14], a correlation with group contributions termed *B* (bulk), *C* (cohesiveness), *D, E,* and *F*; Lai et al. [15], a correlation with group contributions; Stein and Brown [16], a correlation with group contributions; Yalkowsky et al. [17] and Walters et al. [18], a correlation with melting point, molecular symmetry, atom counts, and number of hydrogen bond donors; and Simamora and Yalkowsky [19], a correlation with group contributions.

Mention also must be made of other group contribution methods; that of Joback and Reid [20] has been widely used as a comparison with other methods. For example, Gani and coworkers developed a method that uses both first- and higher-order contributions, and that claims average absolute errors of 5.4° for a data set of 392 diverse compounds [21], and of 5.9° for a diverse set of 1,211 compounds [22], compared with 12.9° and 11.0°, respectively, when using the method of Joback and Reid [20]. Katritzky et al. [23–25] have given overviews of recent QSPR predictions of boiling point.

Banks's [12] method has the advantage of giving a quick and readily calculable boiling point

$$\log T_b \text{ (K)} = 2.98 - 4/\sqrt{MW} \qquad (2)$$

When using Banks's method, Lyman [9] found an average absolute error of 8.8%. Gold and Ogle [26] found an average error for Burnop's method [13] of −1.46%. However, the method of Gold and Ogle of calculating average error is open to question, because the negative value quoted above leads one to suspect that they might have averaged the percentage errors rather than the absolute percentage errors; this could lead to, for example, an error of +6% and one of −6%, averaging to 0% error.

Cramer's [14] *BC(DEF)* group contribution method gave excellent results for compounds in the training set of 114 compounds (standard error *s* = 3.3°). However, Cramer found that as similarity with training set compounds decreased, standard error increased; thus, for compounds containing no more than one structural fragment not found in the training set, the standard error was 38.7°.

The nonlinear group contribution method of Lai et al. [15] gives good results, with a test set of 1,169 compounds yielding an average absolute error of 1.29%. However, it is a complicated method, incorporating a number of correction factors such as decreasing fragment constant values with increasing chain length. The group contribution method of Joback and Reid [20] is simple to use, and their training set of 438 compounds yielded an average absolute error of 12.9° (3.6%).

The group contribution method of Stein and Brown [16] used a very large training set of 4,426 compounds, and thus has very wide applicability. This method is relatively simple to use, and the authors found an average absolute error of 15.5° (3.3%).

The method of Yalkowsky et al. [17] and Walters et al. [18] requires knowledge of a compound's melting point, which thus needs to be determined experimentally or by prediction. This places constraints on the method, and makes it less attractive. The method of Simamora and Yalkowsky [19] is a group contribution method, but is restricted to aromatic compounds. For the 444-compound training set, a standard error of 17.6° was reported; the authors also reported that for the same set of compounds, the method of Joback and Reid [20] gave a standard error of 42°. Krzyzaniak et al. [27] later reported, for a set of 870 diverse aliphatic compounds, $r^2 = 0.999$ and root mean square error = 14.4° when using group contributions and a molecular flexibility term. A comparison with the method of Joback and Reid when using 87 compounds gave a root mean square error of 14.5° compared with 21.0°.

Recent years have seen the publication of a plethora of QSPR methods for the prediction of boiling point, and it is impracticable to cover all of these in a review of this nature. Table 1 lists those from 1996 onwards [28–83]. Notice that many of the studies deal with specific classes of compounds, especially alkanes. Clearly, modeling a property is easier when one is dealing with a single chemical class; however, from the point of view of an environmentalist concerned with a wide range of compounds, the methods of greater interest are those that can adequately model the behavior of diverse data sets.

As far as hydrocarbons are concerned, many of the QSPRs cited in Table 1 are capable of predicting boiling points with a standard error of only a few degrees; arguably the best is the neural network approach of Espinosa et al. [48], who modeled the boiling points of 327 aliphatic hydrocarbons with a combination of topological and physicochemical descriptors in a fuzzy ARTMAP architecture (a synthesis of neural networks and fuzzy logic for categorization and prediction) to yield a standard error of prediction of 1.15°. Also pertinent is the observation that statistical techniques such as multiple linear regression are still used inappropriately in QSPR modeling. For example, Araujo and Morales [43] reported a QSPR containing four topological indices to model the boiling points of nine heptane isomers, in contravention of the rule of Topliss and Costello [84] that the ratio of observations to descriptors in a QSPR should be ≥5, to minimize the risk of chance correlation. Furthermore, at least some of their descriptors (e.g., $^1\chi$ and $^2\chi$) are likely to be highly collinear, with a consequent risk of statistical instability [85]. It is interesting, but not surprising, that almost all the studies of hydrocarbons involve the use of graph theoretical descriptors; such descriptors are very good at modeling molecular branching.

Many recent studies have involved specific chemical classes such as alcohols, amines, and sulfides. Again, it may be noted that most of the studies have employed graph theoretical descriptors, and have standard errors in single figures. Clearly, topological, topochemical, and electrotopological descriptors are powerful tools for the modeling of boiling points.

The boiling point prediction models that have been developed for diverse data sets are grouped together for convenience at the bottom of Table 1. On the whole, the models are quite good, with standard errors of ≤18°. The WHIM descriptors (developed by R. Todeschini, University of Milan, Milan, Italy), which are topochemical in nature, yield [81] a standard error of 16.8° for a relatively small data set (*n* = 69), whereas the CODESSA [23,24] and ADAPT [76,77] software packages, both of which use structural and physicochemical descriptors, give standard errors of 11 to 15°; the use of a neural

Table 1. Recent (1996 onward) quantitative structure–property relationship (QSPR) correlations of normal boiling point

Compounds	Method[a]	Types and number of descriptors	n	r^2	Error[b]	Reference
Alkanes	MLR	Graph theoretical (4)	73	0.996	2.5	[28]
Alkanes	MLR	Graph theoretical (5)	74	0.994	3.7	[29]
Alkanes	MLR	Topological (5)	74	0.986	5.6	[30]
Alkanes	MLR	Topological (2)	74	0.994	3.8	[31]
Alkanes	MLR	Topochemical (10)	125	0.990	5.0	[32]
Alkanes	MLR	Topological (4)	134	0.990	2.7	[33]
Alkanes	NN	Graph theoretical (NG[c])	67	0.999	1.6	[34]
Alkanes	NN	Topological (NG[c])	74	0.990	5.5	[35]
Alkanes	MLR	Quantum chemical (1)	39	0.968	8.7	[36]
Alkanes	MLR	Structural (20)	150	0.998	2.9	[37]
Alkanes	NN	Structural (10)	150	0.997	3.6	[37]
Alkanes	MLR	Topological (5)	38	1.000	0.6	[38]
Alkanes[d]	MLR	Topological (2)	18	0.951	NG[c]	[39]
Alkanes	MLR	Topological (4)	20	1.000	1.0	[40]
n-Alkanes[d]	NLR	Carbon number	28	NG[c]	0.04	[41]
Cycloalkanes[d]	MLR	Topological (4)	209	0.991	3.9	[6]
Alkanes and cycloalkanes	MLR	Distance indices (2)	180	0.978	6.6	[42]
Heptanes	MLR	Topological (2)	9	0.999	2.4	[43]
Octanes	MLR	Topological (1)	18	0.689	3.5	[44]
Nonanes	MLR	Topological (4)	35	0.937	1.6	[45]
Alkenes	NN	Topological (5)	85	NG[c]	2.3%	[46]
PAHs	MLR	WHIM (2)	53	0.982	11.0	[47]
Aliphatic hydrocarbons	NN	Topological and physicochemical (8)	327	1.000	1.2	[48]
Hydrocarbons	MLR	Physicochemical and structural (8)	235	0.998	5.0	[49]
Alkanes and Alcohols	MLR	Polarizability (3)	328	0.981	6.0	[50]
Alkanes, alcohols, and chloroalkanes	NN	Electropological (6)	372	0.941	5.4	[51]
Hydrocarbons	MLR	Topological (NG[c])	153	0.995	4.7	[52]
Aldehydes and ketones	MLR	Topological (12)	72	0.998	3.9	[52]
Haloalkanes	MLR	Structural fragments (4)	58	0.984	7.9	[53]
Haloalkanes	MLR	CODESSA (5)	534	0.980	6.3	[54]
Chlorofluoroethanes	MLR	Structural (6)	53	0.998	2.6	[55]
Alcohols	MLR	Weighted path numbers (3)	58	0.988	3.9	[56]
Alcohols	MLR	Weighted path numbers (3)	58	0.994	2.9	[57]
Alcohols	MLR	Topological (3)	37	0.992	3.1	[57]
Alcohols	MLR	Topochemical (10)	106	0.990	3.2	[58]
Alcohols	MLR	Topological (1)	62	0.903	9.5%	[59]
Alcohols	MLR	Topological (2)	37	0.985	4.4	[59]
Alcohols	MLR	Electrotopological (3)	28	0.926	5.8	[60]
Alcohols	LR	Quantum chemical (1)	15	0.995	5.6	[61]
Alcohols	NN	Topological (NG[c])	50	0.978	7.6	[35]
Alkylnitriles	MLR	Topological (2)	13	0.953	6.1	[62]
Amines	MLR	Topological (2)	34	0.941	14.5%	[59]
Amines	MLR	Electrotopological (4)	21	0.998	3.1	[60]
Chlorobenzenes	MLR	WHIM (1)	13	0.992	5.6	[63]
Chlorosilanes	MLR	CODESSA (8)	50	0.996	6.1	[64, 65]
Chlorosilanes	NN	CODESSA (8)	50	NG[c]	4.1	[64, 65]
Haloaromatics	MLR	WHIM (2)	109	0.966	10.5	[66]
Non–hydrogen-bonded aromatics	MLR	Structural, symmetry, flexibility (13)	191	NG[c]	8.7	[67]
Sulfides	MLR	Topological (2)	21	0.996	1.8	[68]
O and S compounds	MLR	Topological (5)	185	0.980	6.9	[69]
O and S compounds	MLR	CODESSA (5)	185	0.984	6.3	[54]
Acyclic carbonyl compounds	MLR	Topological (5)	200	0.964	6.9	[70]
Esters	MLR	Infrared (9)	41	0.892	33.7	[71]
Diverse	MLR	CODESSA (4)	298	0.973	12.4	[23]
Diverse	MLR	CODESSA (8)	584	0.965	15.5	[24]
Diverse	NN	Structural (25) and topological (1)	135	0.978	11.4	[72]
Diverse	MLR	Physicochemical and topological (4)	63	0.846	23.4	[73]
Diverse	GSA	Physicochemical and topological (12)	268	0.963	12.7	[74]
Diverse	MLR	Atomic contributions (18)	298	0.96	15.5	[75]
Diverse	MLR	Group contributions	1,794	NG[c]	8.1	[22]
Diverse, no N	MLR	ADAPT (10)	248	0.982	11.6	[76]
Diverse, no N	NN	ADAPT (10)	223	NG[c]	9.1	[76]
Diverse	MLR	ADAPT (8)	268	0.980	10.6	[77]
Diverse	NN	ADAPT (8)	241	NG[c]	7.8	[77]
Diverse	MLR	Topochemical (8)	1,023	0.961	18.0	[78]
Diverse	MLR	Topostructural, topochemical, hydrogen bonding (8)	1,015	0.970	15.7	[79]
Diverse	NN	Electrotopological (19)	298	0.995	5.4	[80]
Diverse	MLR	WHIM (5)	69	0.968	16.8	[81]
Diverse	MLR	Physicochemical (6)	152	0.815	NG[c]	[82]
XCH_2CH_2Y	MLR	CODESSA (4)	59	0.913	16.4	[83]

[a] MLR = multiple linear regression; NN = neural network; NLR = nonlinear regression; LR = linear regression; GSA = generalized simulated annealing.

[b] Standard, root mean square, or mean absolute error.

[c] NG = not given.

[d] One example of a number of QSPRs given for different groups of alkanes or other compounds.

network instead of multiple linear regression improves the performance of the ADAPT software considerably, for example from 10.6° to 7.8° [77]. The WHIM algorithm can be downloaded without charge [www.disat.unimib.it/chm/Dragon.htm], and the other two are available commercially (CODES-SA [www.semichem.com] ADAPT [zeus.chem.psu.edu]).

Undoubtedly the best model to date is that developed by Hall and Story [80]. By using atom-type electrotopological state descriptors [60] and a neural network with 19:5:1 architecture, they obtained, for a data set of 298 compounds with a boiling point range of about 430°, a root mean square error of 5.36° and an absolute mean error of 3.93°. This is an outstanding result, and demonstrates the power and versatility of these descriptors.

A number of software packages are available, either commercially or on the Internet, that can predict boiling point. To test their predictive capability, a test set of 100 diverse compounds was selected from the published literature [24,76,80]; boiling points ranged from 187.5 to 560 K. Six packages were tested; the results are listed in Table 2, and are summarized below.

Melting Point, Boiling Point, Vapor Pressure (MPBPVP), version 1.4. This program is available from Syracuse Research Corporation (SRC) [esc.syrres.com] and it and other SRC property prediction programs can now be downloaded without charge from the U.S. Environmental Protection Agency Web site [www.epa.gov/oppt/exposure/docs/episuitedl.htm]. The program accepts simplified molecular input line entry system (SMILES) notation and can run in batch mode, so it is quick and easy to use. The average absolute boiling point prediction error for the 100-compound test set was 13.8°; 14 compounds had errors between 20° and 30°, 3 compounds had errors between 30° and 40°, and 3 compounds had errors > 40° (the worst was cyanogen with an error of 186.2°).

SPARC. This program is available for online access [http://ibmlc2.chem.uga.edu/sparc] from the University of Georgia. It accepts SMILES notation, but each compound must be entered manually. The average absolute boiling point prediction error for the test set was 6.3°; few compounds had errors between 20° and 30°, none had errors between 30° and 40°, and two had errors > 40° (the worst was acetic acid with an error of 45.1°).

Advanced Chemistry Development (ACD) program. This program is available from ACD [www.acdlabs.com]. This package does not accept SMILES, but requires a structural input such as Molfile. Structures can be imported as batch files, but cannot be run in batch mode. The presentation, in both graphical and tabular form, gives boiling points at a range of pressures up to 760 mm Hg. Compounds used in the training set are indicated as having experimentally determined enthalpies of vaporization. These compounds were removed from the test set of 100 compounds so as to give a truer indication of the program's predictive ability. The remaining 54 compounds had an average absolute boiling point prediction error of 1.0°; all but one compound had errors < 5° (the worst was methyl mercaptan with an error of 7.4°).

Molecular Modeling Pro. This program is available from ChemSW [www.chemsw.com]. It accepts SMILES notation, and can be run in batch mode. The average absolute boiling point prediction error for the test set was 21.7°; 24 compounds had errors between 20° and 40°, 9 compounds had errors between 40° and 60°, and 4 compounds had errors > 60° (the

worst was cyanogen with an error of 343.3°, but note that the boiling points of dinitriles are anomalous [62]).

ProPred. This program is available from the Technical University of Denmark [www.capec.kt.dtu.dk]. The program accepts SMILES notation, and can be run in batch mode. It did not calculate a boiling point for formaldehyde. The average absolute boiling point prediction error for the test set was 16.1°; 15 compounds had errors between 20° and 40°, 3 compounds had errors between 40° and 60°, and 5 compounds had errors > 60° (the worst was 1,1-difluoroethene with an error of 163.7°).

ChemOffice. This program is available from CambridgeSoft [www.cambridgesoft.com]. This program accepts SMILES notation and can be run in batch mode. It did not calculate a boiling point for pyrrole. Its boiling point predictions were identical to those from MPBPVP.

VAPOR PRESSURE

Vapor pressure (p_v) is defined as the partial pressure of a compound in equilibrium with its pure liquid or solid at a specified temperature. Vapor pressure clearly is temperature-dependent, as is shown by the Clausius–Clapeyron equation: $dp_v/dT = \Delta H_v/RT^2$, where ΔH_v is the latent heat of vaporization. Vapor pressure is controlled by the same molecular properties that control boiling point (see above).

Vapor pressure, among other factors, controls the residence time of a compound in soil and aqueous milieux, and thus has a profound effect on the ability of a chemical to disperse through the environment. Vapor pressure often is used in the calculation of Henry's law constant from vapor pressure and aqueous solubility. Vapor pressure also has been used to model some types of toxicity [86,87].

Many models for the prediction of vapor pressure are based on the Clausius–Clapeyron equation, and hence permit the estimation of p_v at different temperatures; these have been discussed by Nendza [8] and Sage and Sage [88]. An example is the model developed by Lyman [89]

$$\ln p_v = K_p \ln(RT_b)/0.97[1 - (3 - 2T/T_b)^m/(T/T_b)$$
$$- 2m^*(3 - 2T/T_b)^{m^*-1}\ln T/T_b]$$
$$+ \ln \Delta p_s \quad (3)$$

where p_v is in atmospheres, K_p is a class-specific constant, T_b is the normal boiling point (K), m is a specific factor for liquids ($m = 0.19$) or solids (if $T/T_b > 0.6$, $m = 0.36$; if $0.6 > T/T_b > 0.5$, $m = 0.8$; if $T/T_b < 0.5$, $m = 1.19$), m^* is $-0.2575T/T_m + 0.4133$, T_m is the melting point (K), and $\ln \Delta p_s$ is a correction term for solids only.

Sage and Sage [88] recommend the method of Antoine [90] for liquids that boil below 200°, and the Grain–Watson method [91] for less volatile substances. The former involves an empirical equation, $\ln p/p_b = B[1/(T - C) - 1/(T_b - C)]$, with the constants determined by using a least-squares fitting procedure. However, the Antoine equation is less accurate for lower vapor pressures, because of unreasonable assumptions concerning the variation of enthalpy of vaporization ΔH_b. In such cases the Grain–Watson method is used

$$\ln p/p_b = \Delta H_b/RT_b[1 - (3 - 2T/T_b)^m/(T/T_b)$$
$$- 2m(3 - 2T/T_b)^{m-1}\ln T/T_b] \quad (4)$$

where m is $0.4133 - 0.2575T/T_b$. A correction is applied for solids. Yalkowsky et al. [92] attempted to allow for entropic

Table 2. Experimental and predicted values of boiling point (b.p.) (°C), log(vapor pressure) (v.p.) (Pa), and melting point (m.p.) (°C) for the 100-compound test set[a]

Compound	b.p. (exp.)	b.p. (MPBPVP)	b.p. (SPARC)	b.p. (ACD)	b.p. (MMP)	b.p. (ProPred)[b]	b.p. (Chem-Office)	log v.p. (exp.)	log v.p. (MPBPVP)	log v.p. (SPARC)	log v.p. (ACD)	log v.p. (MMP)	m.p. (exp.)	m.p. (MPBPVP)	m.p. (ProPred)	m.p. (Chem-Office)
Acetic acid	118.0	122.3	163.1	117.1	160.3	124.0	122.4	3.318	3.274	2.911	3.254	2.274	16.6	-21.3	35.3	-0.7
Acetone	56.3	44.8	35.7	46.5	65.5	33.6	44.9	4.488	4.695	4.460	4.664	4.301	-94.3	-93.6	-82.0	-100.1
Acetonitrile	81.7	71.8	72.4	63.5	140.5	32.4	71.9	4.084	4.211	4.070	4.355	2.627	-48.0	-83.8	-69.8	-96.3
Acetophenone	202.0	189.8	201.7	202.0	141.6	203.0	189.9	1.722	1.902	1.663	1.597	2.794	19.6	-9.9	3.8	-17.3
Aniline	184.1	184.0	189.8	184.5	183.7	190.0	184.1	1.955	2.025	1.845	1.990	1.717	-6.2	-6.2	12.1	-6.5
Benzene	80.1	102.2	77.3	78.9	86.5	85.9	102.3	4.090	3.656	4.138	4.127	3.983	5.5	-77.9	-87.7	-102.3
Benzyl chloride	179.5	184.2	168.1	179.4	156.5	180.0	184.3	2.240	2.021	2.360	2.249	2.553	-39.0	-27.2	-24.7	-48.6
Bicyclohexyl	239.1	222.8	220.6	239.1	220.1	236.0	222.9	1.156	1.360	0.758	0.922	1.246	4.0	-8.5	-1.4	-33.8
1-Bromopropane	71.0	77.6	72.6	71.8	75.6	71.3	77.7	4.266	4.150	4.231	4.247	4.181	-110.0	-79.1	-113.0	-90.2
Bromotrichloromethane	104.9	120.1	102.9	103.5	104.7	138.0	120.1	3.716	3.318	3.800	3.699	3.675	-6.0	-31.9	-3.8	-20.6
1-Butanal	74.9	84.8	81.3	77.7	86.4	76.7	84.9	4.175	3.978	4.066	4.105	3.926	-96.0	-80.2	-60.9	-96.8
n-Butane	-0.4	19.6	-0.9	-1.6	13.2	-19.4	19.7	5.387	5.088	5.394	NP	5.006	-138.4	-120.3	-179.0	-138.8
2,3-Butanediol	180.8	156.8	154.8	180.7	193.6	206.0	156.8	1.385	1.977	1.932	1.543	0.786	25.0	-34.4	1.1	-47.1
n-Butylacetate	126	125.8	131.3	126.7	136.0	130.0	125.9	3.174	3.208	3.126	3.185	2.891	-106.2	-56.8	-83.6	-73.9
n-Butylethyl ether	92.3	97.1	96.2	92.3	90.4	92.8	97.2	3.867	3.812	3.797	3.893	3.591	-103.0	-75.3	-110.2	-94.0
tert-Butylmercaptan	64.3	83.1	67.4	67.1	114.3	68.7	83.2	4.383	4.009	4.353	4.327	3.403	1.0	-82.3	-4.1	-99.9
n-Butyronitrile	117.7	120.2	108.4	117.4	143.9	131.0	120.2	3.415	3.315	3.474	3.366	2.593	-112.0	-58.4	-87.4	-73.8
Chlorodifluoromethane	-40.8	-15.4	-40.8	-38.1	-40.8	-35.8	-15.3	6.016	5.616	5.915	NP	5.006	-146.0	-139.4	NP	-156.5
Chloroform	61.2	91.2	62.4	60.3	61.7	78.6	91.3	4.418	3.862	4.407	4.426	4.417	-63.7	-78.9	-66.3	-97.8
2-Chloropropene	22.7	23.7	21.2	22.6	50.4	20.2	23.7	5.020	5.026	5.081	5.039	4.600	-138.0	-117.6	-106.1	-135.8
Chlorotrifluoromethane	-81.4	-39.4	-77.4	-77.9	-109.2	-62.8	-39.3	6.551	5.927	6.261	NP	5.006	-181.0	-137.3	-112.0	-138.5
3-Cresol	202.3	190.8	191.7	202.3	193.3	197.0	190.9	1.267	1.615	1.361	1.439	0.857	11.5	15.7	31.2	33.2
Cumene	152.5	157.0	150.6	152.5	164.4	152.0	157.1	2.784	2.586	2.913	2.774	2.391	-96.0	-46.3	-73.8	-71.0
Cyanogen	-21.1	165.1	-61.2	-21.1	322.2	133.0	165.2	5.758	2.419	6.197	NP	-2.668	-34.4	-24.1	4.0	-31.3
Cyclohexanol	160.9	161.7	153.6	159.6	152.5	159.0	161.8	1.942	1.968	2.237	2.063	1.932	24.0	-33.4	14.7	-48.0
Cyclohexanone	155.8	155.0	147.6	155.8	131.4	160.0	155.1	2.742	2.740	2.808	2.597	2.988	-47.0	-29.6	-7.4	-36.4
Cyclopentane	49.3	62.5	50.8	49.3	43.5	41.8	62.6	4.600	4.406	4.603	4.620	4.716	-94.0	-94.5	-118.7	-112.4
1-Decanol	231.0	238.6	232.0	227.8	212.1	236.0	238.7	0.056	-0.046	0.042	0.322	0.484	-15.0	7.9	-15.2	-10.3
1,2-Dichlorobenzene	180.5	174.7	177.2	180.5	175.6	170.0	174.8	2.257	2.222	2.286	2.222	2.156	NA	-14.3	-32.0	-17.4
1,2-Dichlorobutane	124.0	124.7	114.1	122.8	117.7	134.0	124.8	3.350	3.229	3.401	3.341	3.354	-97.4	-67.2	-72.6	-93.9
1,1-Dichloroethane	57.3	63.0	62.8	57.3	18.0	45.8	63.1	4.481	4.368	4.390	4.488	5.006	-35.3	-96.4	-100.9	-116.5
1,2-Dichloroethane	83.5	104.7	77.2	83.5	66.7	96.9	104.8	4.027	3.611	4.136	4.049	4.332	-84.0	-76.8	-74.3	-101.5
1,3-Diethylbenzene	181.2	190.8	183.4	181.7	181.8	191.0	190.8	2.200	1.882	2.280	2.194	2.027	-116.3	-17.0	-95.9	-32.2
Diethyl ether	34.5	47.0	41.9	33.2	35.8	21.0	47.1	4.856	4.659	4.771	4.878	4.784	-144	-101.2	-137.1	-116.5
1,1-Difluoroethene	-85.6	-40.4	-84.9	-85.7	-113.8	78.1	-40.3	6.605	5.894	6.409	NP	5.006	5.0	-156.3	-125.9	-175.8
Diiodomethane	182.1	173.3	188.4	182.0	200.9	162.0	173.0	2.380	2.380	2.312	2.187	1.654	-37	-55.9	-63.6	-56.5
Dimethylamine	6.9	16.1	8.7	6.2	-10.3	-21.3	16.1	5.308	5.146	5.336	NP	5.006	2.5	-106.2	-106.3	-108.6
N,N-Dimethylaniline	193.6	169.4	179.8	193.6	157.4	192.0	169.5	1.988	2.332	2.065	1.787	2.338	-157.0	-24.6	-6.9	-34.8
2,3-Dimethyl-1-butene	55.7	47.3	59.7	55.9	85.8	47.6	47.4	4.510	4.655	4.539	4.512	3.996	-84.7	-116.3	-144.0	-146.9
Dimethyl disulfide	109.8	113.7	98.7	109.8	162.4	125.0	113.7	3.583	3.441	3.773	3.583	2.380	NA	-92.0	-91.4	-92.5
2,3-Dimethylpentane	89.8	66.4	86.6	89.5	107.9	85.4	66.4	3.962	4.342	4.025	3.944	3.618	-98.0	-104.7	-129.8	-135.0
Dimethyl sulfide	37.4	42.5	27.2	29.6	61.3	26.2	42.6	4.810	4.718	4.997	4.933	4.405	11.8	-107.7	-143.4	-126.9
1,4-Dioxane	101.0	102.8	98.4	102.9	60.0	99.8	102.8	3.707	3.710	3.673	3.707	4.265	-114.1	-63.6	-34.5	-74.0
Ethanol	78.3	65.1	80.8	72.7	77.7	42.5	65.2	3.899	4.189	3.859	4.040	3.870	-135.4	-87.8	-78.8	-100.5
Ethyl n-butyrate	121.5	125.8	127.7	122.5	136.0	130.0	125.9	3.349	3.208	3.184	3.266	2.891	-143.2	-56.8	-87.3	-73.9
Ethylfluoride	-37.7	-33.2	-35.9	-36.5	-87.6	-82.8	-33.1	5.963	5.812	5.878	NP	5.006	-76.0	-146.7	-153.2	-160.7
2-Ethyl-1-hexanol	184.7	188.5	180.0	184.6	180.1	193.0	188.6	1.283	1.292	1.529	1.439	1.275	-62	-25.5	-33.9	-47.9
4-Ethyltoluene	162.1	170.0	166.2	161.8	160.3	171.0	170.1	2.600	2.318	2.634	2.590	2.474	-116	-28.7	-31.8	-43.5
Ethyl vinyl ether	35.6	45.1	42.6	32.5	44.7	28.9	45.2	4.837	4.691	4.797	4.887	4.596	-98.0	-102.6	-128.8	-118.3
Fluorobenzene	85.0	96.7	86.5	84.8	84.2	84.0	96.8	4.012	3.760	4.060	4.024	4.024	-41.2	-73.0	-72.1	-89.2
Formaldehyde	-19.1	9.5	-14.2	-19.5	-19.5	NP	9.6	5.715	5.249	5.814	NP	5.006	-118.0	-110.9	NP	-114.2

Table 2. Continued

Compound	b.p. (exp.)	b.p. (MPBPVP)	b.p. (SPARC)	b.p. (ACD)	b.p. (MMP)	b.p. (ProPred)[b]	b.p. (ChemOffice)	log v.p. (exp.)	log v.p. (MPBPVP)	log v.p. (SPARC)	log v.p. (ACD)	log v.p. (MMP)	m.p. (exp.)	m.p. (MPBPVP)	m.p. (ProPred)	m.p. (ChemOffice)
Furan	31.4	59.3	25.0	31.4	59.0	47.4	78.1	4.903	4.432	4.959	4.907	4.288	−85.6	−87.1	−77.4	−95.5
n-Heptane	98.5	96.2	98.5	98.8	100.5	96.7	96.2	3.783	3.828	3.847	3.779	3.748	−90.6	−81.0	−129.5	−105.0
n-Hexadecane	286.9	277.1	291.0	286.6	280.1	274.0	277.2	−0.700	0.162	−0.835	−0.222	−1.057	18.2	22.5	−38.4	−3.5
1-Hexadecene	284.9	275.9	292.9	284.9	284.9	273.0	276.0	−0.578	0.189	−0.442	−0.155	−1.198	4.0	21.3	−34.9	−5.3
1-Hexane	63.5	69.7	68.6	62.9	81.6	61.5	69.7	4.370	4.286	4.379	4.397	4.069	−139.8	−95.3	−137.1	−118.0
n-Hexanoic acid	205.8	207.8	209.8	204.7	184.6	211.0	207.8	0.763	1.465	0.872	1.338	1.846	−3.0	26.2	52.6	44.4
3-Hexanone	123.5	118.8	118.6	125.6	129.0	118.0	118.9	3.268	3.422	3.370	3.204	3.036	NA	−55.1	−54.1	−66.3
Isobutene	−6.8	10.2	−3.4	−4.9	22.0	−18.9	10.3	5.482	5.229	5.436	NP	5.006	−140.3	−130.9	−152.8	−154.5
Isobutyl acetate	116.7	111.7	123.9	116.7	140.2	119.0	111.8	3.376	3.477	3.228	3.378	2.803	−99.0	−68.4	−86.2	−88.9
Isobutyl formate	98.1	90.0	94.1	99.5	119.7	98.0	119.3	3.738	3.884	3.788	3.706	3.212	−95	−72.2	−92.3	−78.3
Isopropyl mercaptan	52.6	68.1	57.2	55.8	85.7	50.8	68.2	4.567	4.277	4.492	4.512	3.988	−131.0	−101.0	−139.3	−128.6
Methanol	64.7	39.4	57.3	48.1	65.1	0.3	58.5	4.226	4.723	4.397	4.549	4.218	−98.0	−101.0	−89.1	−111.7
Methyl acetate	57.0	52.6	41.6	44.1	73.0	33.8	52.7	4.456	4.547	4.425	4.688	4.164	−98.0	−95.1	−118.1	−107.7
Methylamine	−6.3	10.2	−9.6	−21.1	−8.1	−37.7	10.2	5.548	5.239	5.590	NP	5.006	−38.0	−98.3	−67.1	−89.3
Methyl n-butyl sulfide	123.5	116.8	116.8	122.7	132.6	125.0	116.8	3.318	3.383	3.409	3.344	3.040	NA	−69.1	−103.2	−93.1
Methyl ethyl ketone	79.7	70.4	75.6	75.7	86.4	69.4	70.4	4.090	4.274	4.106	4.181	3.926	−86.3	−80.5	−75.9	−88.8
Methyl fluoride	−78.3	−62.2	−70.5	−78.2	−78.4	−155.0	−62.2	6.582	6.073	6.471	NP	5.006	−141.8	−160.8	−176.5	−172.0
Methyl iodide	42.5	64.1	38.7	40.4	42.6	36.9	64.1	4.732	4.380	4.799	4.763	4.730	−66.5	−95.2	−130.0	−114.5
Methyl mercaptan	6.0	32.0	9.7	−1.4	28.7	−15.6	32.1	5.304	4.892	5.309	NP	4.948	−121.0	−115.3	−198.8	−136.1
4-Methylpyridine	145.4	136.4	141.8	144.3	108.3	142.0	136.5	2.885	2.999	2.990	2.933	3.597	2.4	−25.5	14.3	−18.3
Methylstyrene	165.5	162.6	170.1	162.5	168.8	167.0	162.6	2.560	2.473	2.524	2.576	2.299	−24.0	−45.0	−37.6	−71.7
1-Octanol	195.2	200.7	196.2	194.7	176.4	201.0	200.7	1.024	0.976	1.044	1.190	1.375	−15.0	−14.5	−28.8	−32.9
Pentachloroethane	159.9	152.2	156.6	161.1	144.2	168.0	152.3	2.670	2.684	2.603	2.600	2.828	−29.0	−23.8	13.2	−24.3
2-Pentanone	102.4	95.0	101.0	102.7	107.8	101.0	95.1	3.676	3.848	3.740	3.709	3.536	−78.0	−67.7	−66.5	−77.6
Phenylhydrazine	243.6	201.2	251.7	243.1	169.2	219.0	201.3	0.535	1.615	0.044	0.643	2.049	19.0	25.2	29.1	46.1
Piperidine	106.3	127.9	112.4	106.4	74.0	136.0	128.0	3.632	3.166	3.430	3.578	4.153	−13.0	−24.7	−1.0	−10.8
Propane	−42.0	−7.8	−44.3	−42.7	−23.4	−77.1	−7.7	5.979	5.485	5.967	NP	5.006	−187.7	−133.9	−200.0	−150.0
1-Propanol	97.3	90.0	94.5	95.8	86.9	78.0	90.0	3.449	3.651	3.592	3.548	3.662	−126.0	−75.0	−69.2	−89.2
Propene	−47.7	−9.8	−45.4	−47.7	−13.7	−74.1	−9.8	6.060	5.512	5.980	NP	5.006	−185.2	−135.4	−176.9	−151.8
Propionic acid	141.2	145.0	166.3	141.7	154.6	149.0	145.1	2.695	2.828	2.310	2.749	2.425	−21.5	−9.0	39.9	10.6
n-Propylcyclohexane	156.8	152.6	157.7	155.8	153.2	159.0	152.6	2.747	2.787	2.590	2.708	2.645	−95.0	−49.6	−83.6	−75.0
Pyridine	115.3	113.4	115.1	115.3	80.2	110.0	113.4	3.443	3.447	3.502	3.486	4.094	−41.6	−44.5	−36.6	−42.0
Pyrrole	129.8	126.3	123.4	129.8	62.2	101.0	NP	3.040	3.197	2.944	3.212	4.372	−23.0	−44.8	−12.2	NP
Quinoline	237.7	240.2	231.1	234.2	190.1	235.0	240.3	1.047	0.654	1.002	1.037	1.602	−15.0	37.6	−15.3	48.3
Styrene	145.2	146.7	150.1	145.2	146.7	147.0	146.7	2.912	2.795	2.945	2.914	2.754	−30.6	−48.3	−61.0	−69.0
Tetrachloromethane	76.7	80.4	75.9	76.1	76.5	103.0	80.4	4.182	4.059	4.213	4.174	4.166	−22.9	−58.4	−9.0	−50.5
Tetrahydrofuran	64.9	69.4	63.2	68.3	35.5	52.4	69.4	4.333	4.292	4.371	4.305	4.791	−108.4	−84.9	−81.7	−97.0
Thiophene	84.2	113.9	87.3	84.2	140.9	107.0	114.0	4.022	3.437	4.102	4.041	2.872	−38.2	−45.7	−69.1	−38.6
1,1,1-Trifluoroethane	−47.3	−56.5	−70.8	−47.6	−119.2	−114.0	−56.4	6.101	6.043	6.235	NP	5.006	−107.0	−151.7	−158.3	−157.1
Trifluoromethane	−82.1	−74.3	−84.2	−84.4	−82.3	−23.2	−74.2	6.650	6.113	6.508	NP	5.006	−155.2	−171.2	−92.7	−185.8
1,2,4-Trimethylbenzene	169.4	170.0	167.7	170.8	160.3	162.0	170.0	2.458	2.321	2.495	2.408	2.474	−43.8	−22.5	−54.3	−30.9
1,3,5-Trimethylbenzene	164.8	170.0	162.0	166.7	160.3	162.0	170.0	2.527	2.321	2.610	2.494	2.474	−44.7	−22.5	−54.8	−30.9
2,2,3-Trimethylbutane	81.0	56.3	80.0	81.8	111.7	79.3	56.4	4.135	4.509	4.171	4.075	3.470	−25.0	−99.0	−121.4	−117.5
2,2,4-Trimethylpentane	99.3	81.4	99.9	98.8	135.5	93.3	81.5	3.818	4.085	3.858	3.779	3.001	−107.0	−86.0	−93.7	−106.3
Valeronitrile	141.4	143.0	130.8	141.7	156.4	156.0	143.0	2.988	2.869	3.086	2.885	2.336	−96.0	−46.1	−77.3	−62.5
Vinyl bromide	15.8	50.3	21.2	15.8	15.8	15.9	50.4	5.130	4.605	5.084	NP	5.006	−139.5	−93.5	−137.8	−103.3
Vinyl fluoride	−72.2	−35.3	−62.1	−73.5	−72.2	−72.0	−35.2	6.440	5.838	6.173	NP	5.006	−160.5	−148.2	−160.5	−162.5
2-Xylene	144.5	148.3	142.2	146.0	137.5	137.0	148.4	2.948	2.762	3.049	2.898	2.940	−25.2	−40.7	−79.4	−54.7
2,4-Xylenol	211.0	210.7	206.2	211.0	207.5	211.0	210.7	1.153	1.106	1.035	1.253	0.537	27.5	25.4	43.3	57.0

[a] exp. = experimental; NA = not available; NP = not predicted.
[b] ProPred, Technical University of Denmark, Lyngby, Denmark.

Table 3. Recent (1990 onward) quantitative structure–property relationship (QSPR) correlations of vapor pressure

Compounds	Method[a]	Types and number of descriptors[b]	n	r^2	Error[c]	Reference
Hydrocarbons	NN	Topological, MW T (7)	274	0.975	0.035	[106]
Hydrocarbons and halohydrocarbons	MLR	ADAPT (7)	352	0.983	0.186	[107]
Hydrocarbons and halohydrocarbons	NN	ADAPT (7)	270	NG[d]	0.163	[107]
Hydrocarbons and halohydrocarbons	NN	Structural, topological, T (25)	1,200	0.995	0.08	[108]
Fluorinated hetero-hydrocarbons	MLR	Physicochemical, topological (6)	16	NG[d]	NG[d]	[109]
Polycyclic aromatics	MLR	Topological (2)	30	0.89	0.97	[110]
Hexanes	MLR	Structural, T (2)	312	0.978	0.14	[111]
Cyclohexanes	MLR	Structural, T (2)	235	0.988	0.11	[111]
Benzenes[e]	LR	Structural (1)	65	0.951	0.13	[112]
Primary amines	LR	Structural (1)	12	0.955	0.49	[113]
Non–hydrogen-bonded aromatics	MLR	Structural, symmetry, flexibility (13)	73	NG[d]	0.38	[67]
Esters	PLS	Infrared (10)	41	0.933	0.627	[71]
Diverse	MLR	Group contributions (42)	342	NG[d]	5.0%	[103]
Diverse	MLR	CODESSA (5)	411	0.949	0.331	[114]
Diverse	MLR	α^f, structural (7)	479	0.960	0.534	[115]
Diverse	NN	α^f, structural (7)	479	0.961	0.522	[115]
Diverse	MAM	Autocorrelation (1)	366	0.978	0.14	[116]
Diverse	MLR	Atomic contributions (31 PCs[g])	1,771	0.88	2.1	[75]
Diverse	MLR	Group contributions (81), hydrogen-bonding (2), and melting point	1,410	0.98	0.36	[104]
Diverse	NN	ADAPT (8)	290	NG[d]	0.26	[117]
Diverse	NN	Quantum mechanical (10)	551	NG[d]	0.37	[118]
Diverse	MLR	Topostructural, topochemical	476	0.843	0.29	[119]
Diverse	MLR	Topostructural, hydrogen-bonding (10), topochemical, hydrogen-bonding (12)	469	0.913	0.22	[120]
Diverse	KNN	Electrotopological (ca. 15 PCs)	1,676	0.733	0.526	[121]
Diverse	MLR	Solvatochromic (3)	53	0.960	NG[d]	[100]
Diverse	MLR	UNIFAC[h] (3)	118	0.903	NG[d]	[101]

[a] NN = neural network; MLR = multiple linear regression; LR = linear regression; MAM = multifunctional autocorrelation method; PLS = partial least squares; KNN = K-nearest neighbor.
[b] MW = molecular weight; T = temperature.
[c] Log_{10} units; standard, root mean square, or mean absolute error.
[d] NG = not given.
[e] One example of several QSPRs for different series of compounds.
[f] Polarizability.
[g] PC = principal component.
[h] UNIFAC = UNIQUAC functional-group activity coefficient.

effects by incorporating terms for molecular symmetry and flexibility.

Although such methods seem to give reasonable predictions, they give the appearance at least of being cumbersome, and they throw no light on the factors that control vapor pressure. Müller and Klein [1] commented in 1993 that estimation methods for vapor pressure needed further development. These equation-of-state methods, many of which require the input of experimental values of one or more properties, will not be discussed further here.

In recent years, attempts have been made to develop QSPRs for vapor pressure prediction, and those from 1990 onward are summarized in Table 3. Note that with few exceptions these methods predict vapor pressure at a single temperature (298 K).

Probably the earliest attempt to model vapor pressure by a QSPR method was that of Prausnitz and coworkers [93–95]. They developed a group contribution approach for 67 hydrocarbons that allowed them to predict vapor pressures over a range of temperatures. Burkhard [96] later extended their approach to include halogenated aromatic hydrocarbons, and found an average prediction error of <20%. Burkhard et al. [97] compared 11 methods for predicting the vapor pressures of 15 polychlorinated biphenyls (PCBs); for those methods that required only calculated properties as input, correlations with MW, molecular surface area, and molecular connectivity

indices were all found to give good results, although correlation coefficients and standard errors were not given. Hawker [98] and Rouvray and Tatong [99] correlated the vapor pressure of PCBs with planar total surface area and topological indices, respectively. Other measures of molecular size would also probably model the vapor pressure of PCBs; for example, Rordorf [100] found that vapor pressure could be correlated with the number of chlorine atoms in dioxins.

Banerjee et al. [101] found that the Kamlet–Taft solvatochromic [102] and UNIFAC (UNIQUAC functional-group activity coefficient, where UNIQUAC is an acronym for universal quasi chemical) [103] descriptors modeled vapor pressure well for a small data set of 53 diverse compounds. The next study to be published was that of Tu [104], who devised a group contribution method. Tu used 5,359 experimental vapor pressures for 342 organic compounds, and generated group contributions for 42 groups. His equation has the form

$$\log pM = \left[\sum N_i(a_i + b_i/T - c_i \ln T - d_i T) \right] + Q \quad (5)$$

where M is molecular weight; N_i is the number of group i in the molecule; a_i, b_i, c_i, and d_i are group constants of group i, and Q is a specific compound correction.

For a test set of 336 organic compounds with 5,287 data points, the mean absolute percentage error was 5.0%. Other

group and atomic contribution methods have been developed by Simmons [105] and Labute [75].

Most of the methods listed in Table 3 [106–121] have involved the use of quite large data sets with a wide range of vapor pressures, and most have quite low standard errors, even when the data set comprises structurally very diverse chemicals. Excluding the study of Labute [75], the average standard error of those studies involving diverse data sets is 0.36 log units, or a factor of 2.3, which is excellent considering the range of vapor pressures involved (often over eight or more orders of magnitude). The high standard error in Labute's study is disappointing, but it must be remembered that his was the largest data set, with 1,771 compounds; nevertheless, atomic contributions alone do not seem to model vapor pressure well.

The autocorrelation method of Chastrette and Crétin [116] yielded a very low standard error (0.14), and this method clearly needs further investigation with an expanded data set. Three other studies on large diverse data sets are worthy of note because of their low standard errors, namely that of Basak et al. [119], who used topological, topostructural, and H-bonding descriptors; the work of McClelland and Jurs [117], who used a neural network approach with ADAPT descriptors; and the study of Beck et al. [118], who used quantum mechanical descriptors. It is interesting (and comforting) to see that good modeling of vapor pressure is not the prerogative of one type of molecular descriptor only.

Entropic factors clearly must play a role in vapor pressure, because transfer from the liquid or solid state to the vapor state increases disorder in a system. Jain and Yalkowsky [67] used a symmetry number that accounts for this in part. It would be useful to explore whether other entropic terms (such as molecular flexibility), in combination with either group contributions or structural, physicochemical, or topological descriptors, would improve predictive ability.

Delle Site [122] reviewed a number of vapor pressure prediction methods, including QSPR methods and those based on the Clausius–Clapeyron equation. Although the latter generally allow prediction of vapor pressures over a range of temperatures, it must be remembered that these methods require the input of at least one experimental variable (e.g., melting point). However, Table 3 shows that three QSPR approaches [106,108,111] allow vapor pressure prediction over a temperature range. Hopefully, this aspect of vapor pressure prediction will be investigated more fully in the near future.

Most of the software packages that predict boiling point (see above) also predict vapor pressure. The results from the 100-compound test set (with p_v in pascals [Pa] at 25°C) are listed in Table 2, and are summarized below.

MPBPVP from Syracuse Research Corporation. This software reports p_v in mm Hg. The average absolute prediction error on log p_v (Pa) was 0.285; 18 compounds had errors between 0.4 and 0.6, 4 compounds had errors between 0.6 and 1.0, and 2 compounds had errors > 1.0 (the worst was cyanogen with an error of 3.339).

SPARC from the University of Georgia. This software reports log p_v in atmospheres (atm.). The average absolute prediction error on log p_v (Pa) was 0.105; two compounds had errors between 0.3 and 0.4, and four compounds had errors > 0.4 (the worst was 2,3-butanediol with an error of 0.547).

ACD from Advanced Chemistry Developments. This software reports p_v in a choice of units, over a wide temperature range. For compounds with normal boiling points more than a few degrees below 25°C, vapor pressure at 25°C was not

predicted; this was the case for 18 of the test set compounds. This, together with the removal of test set compounds that were also in the ACD training set, meant that only 42 of the test set compounds could be used to assess the predictivity of the software. The average absolute prediction error on log p_v (Pa) was 0.107; no compounds had errors between 0.3 and 0.4, and three compounds had errors > 0.4 (the worst was *n*-hexanoic acid with an error of 0.575).

Molecular Modeling Pro from ChemSW. This software reports p_v in torr (mm Hg). The average absolute prediction error on log p_v (Pa) was 0.573; 15 compounds had errors between 0.4 and 0.6, 15 compounds had errors between 0.6 and 1.0, and 17 compounds had errors > 1.0 (the worst was cyanogen with an error of 8.426).

MELTING POINT

The melting point is the temperature at which a solid fuses to become a liquid; the freezing point is the temperature at which a liquid solidifies, and for all practical purposes the two can be considered to be identical. Melting points are almost invariably reported as being measured at atmospheric pressure. As has been shown above, a measure or estimate of melting point is required in some methods for the prediction of boiling point and vapor pressure. New chemical entities submitted to regulatory agencies may have no reported melting point, and it could be more convenient to make an estimate of that property than to request the manufacturer to supply it. It is also useful to know a compound's melting point even before synthesis, in order to be aware of potential handling and disposal problems. Finally, as mentioned above, melting point is useful in one method for the estimation of aqueous solubility [2].

Melting point prediction has a long history, starting as long ago as 1884 [4], although much of the early work involved homologous series. Horvath [7], Dearden [123], and Tesconi and Yalkowsky [124] all have reviewed the prediction of melting point. Melting point is generally acknowledged to be difficult to predict, because a number of factors that control it are not easy to quantify. Melting occurs when the forces of thermal agitation overcome the interactions holding the solid crystal together; these interactions can include ionic, polar, and hydrogen bonding forces. Crystal packing also plays a great part, with symmetrical molecules showing higher melting points. An example is given by anthracene, a symmetrical tricyclic aromatic, which melts at 217°, whereas its asymmetrical isomer phenanthrene melts at 100°. In addition, entropic effects play a major role in melting, with positional, expansional, rotational, and conformational entropy all contributing to the melting point [125–127]. Clearly, the prediction of melting points for diverse data sets should take account of all these factors.

The QSPR developed in 1884 by Mills [4] gave outstandingly good predictions for the melting points of members of homologous series, and over the years many more such QSPRs have been reported [123]. Table 4 lists those published from 1990 onward [128–149]; the reader is referred to the comprehensive review by Dearden [123] for earlier QSPR studies of melting point.

Of course, much work has been done with hydrocarbons. Somayajulu [150] developed an equation based on chain length

$$\ln(T_m^\infty - T_m) = a - bn^{1/25} \qquad (6)$$

where T_m^∞ is the melting point (K) of the compound of infinite chain length, and *a* and *b* are series constants. Similar equa-

Table 4. Recent (1990 onward) quantitative structure–property relationship (QSPR) correlations of melting point

Compounds[a]	Method[b]	Types and number of descriptors	n	r^2	Error[c]	Reference
Alkanes	MLR	Topological (6)	17	0.895	13.6	[128]
Alkanes	MLR	Topological (7)	17	0.901	14.0	[129]
Alkanes	MLR	Topological (2)	17	0.693	20.0	[130]
Alkanes	NN	Structural (8)	150	0.956	8.1	[131]
Hydrocarbons	MLR	Structural (NG[d])	307	NG[d]	7.4%	[132]
PAHs	MLR	WHIM (4)	79	0.887	35.1	[47]
PAHs	MLR	WHIM (6)	79	0.895	NG[d]	[133]
Chlorobenzenes	MLR	WHIM (2)	13	0.934	18.1	[134]
Halomethanes	LR	Physicochemical (1)	30	0.676	47.8	[135]
Halobenzenes, halotoluenes	MLR	WHIM (6)	92	0.862	22.3	[66]
Halobenzenes	LR	Topological (1 PC[e])	25	0.87	21.0	[136]
PCBs	MLR	Structural and physicochemical (5)	58	0.828	22.1	[126]
PCBs	MLR	WHIM (4)	82	0.820	21.3	[137]
Alcohols	LR	Quantum chemical (1)	15	0.947	14.9	[61]
Anilines	MLR	Physicochemical (6)	42	0.897	23.7	[138]
Aldehydes	MLR	Topological (5)	72	0.838	NG[d]	[139]
Amines	MLR	Topological (5)	54	0.795	NG[d]	[139]
Ketones	MLR	Topological (5)	52	0.865	NG[d]	[139]
Caffeines	MLR	Topological (6)	12	0.964	20.5	[129]
Amino acids	MLR	Topological (2)	20	0.760	22.5	[128]
Amino acids	MLR	Topological (2)	20	0.929	12.7	[140]
Pyridines and piperidines	MLR	CODESSA (6)	141	0.831	NG[d]	[141]
Pyrazolinones	MLR	Topological (2)	17	0.891	15.0	[142]
Benzenes	MLR	CODESSA (9)	443	0.837	30.2	[143]
Aromatics	MLR	Group contributions (36) and CFs[f] (5)	1,690	NG[d]	37.5	[19]
Aromatics (rigid)	MLR	Structural (22) and symmetry	1,181	0.991	36.6	[144]
Aromatics, non–hydrogen-bonded	MLR	Structural (11) and symmetry	338	NG[d]	23.1	[67]
Aliphatics, non–hydrogen-bonded	MLR	Group contributions (24) and CFs[f] (8)	596	0.977	34.3	[27]
Aliphatics	MLR	Group contributions (46) CFs[f] (9) and structural	1,040	NG[d]	34.4	[145]
Diverse, rigid non–hydrogen-bonded	MLR	Structural (11) and symmetry	520	0.988	39.2	[146]
Diverse, non–hydrogen-bonded	MLR	Structural (5) and boiling point	979	0.808	35.5	[17]
Diverse[g]	MLR	Structural (11)	178	0.937	24.0	[147]
Diverse[h]	MLR	Structural (11)	366	0.919	17.9	[148]
Diverse	MLR	WHIM (6)	94	0.834	32.8	[81]
Diverse	MLR	Topological and physicochemical (4)	62	0.856	18.0	[73]
Diverse	MLR	Group contributions (15) and structural (6)	1,310	NG[d]	8.2	[149]

[a] PAH = polycyclic aromatic hydrocarbon; PCB = polychlorinated biphenyl.
[b] MLR = multiple linear regression; NN = neural network; LR = linear regression.
[c] Standard, root mean square, or mean absolute error.
[d] NG = not given.
[e] PC = principal component.
[f] CF = correction factor.
[g] Ethenes, benzenes, and naphthalenes with diverse substituents.
[h] Alkanes with diverse substituents.

tions were later proposed by Riazi and Al-Sahhaf [151] and Marano and Holder [41]. Cherqaoui et al. [131] used a neural network approach to predict the melting points of all 150 alkanes up to C_{10}; they obtained $r^2 = 0.956$ and $s = 8.1°$. Pogliani [128–130] used modified connectivity indices with a small series of alkanes, and Todeschini and Gramatica [47] and Todeschini et al. [133] applied their new WHIM descriptors to model melting points of polyaromatic hydrocarbons.

Within congeneric series (i.e., a common core with diverse substituents), Dearden and Rahman [152] obtained a standard error of 20.5° for a series of 41 substituted anilines. Katritzky and Gordeeva [139] found reasonable correlations ($r^2 \geq 0.795$) for series of 27 aldehydes, 48 amines, and 30 ketones separately (*s* not given). Murugan et al. [141] obtained $s = 36.1°$ for a set of 141 substituted pyridines, and Medić-Šarić et al. [142] correlated the melting points of 17 pyrazolinones with molecular connectivities with $r^2 = 0.891$ and $s = 15.0°$. Somayajulu [150] used Equation 6 to predict melting points of homologous series of haloalkanes, alkanols, alkanals, alkanones, alkanoic acids and esters, dialkyl ethers, thiols, alkyl-

amines, and alkanenitriles; standard deviations were generally better than 1°.

Charton and Charton [147,148] used their intermolecular force equation to model the melting points of diversely substituted alkanes, ethenes, benzenes, and naphthalenes with good results (see Table 4). Pogliani [140], again by using his modified connectivity indices, correlated melting points of caffeine derivatives with $r^2 = 0.964$ and $s = 20.5°$, and of amino acids with $r^2 = 0.929$ and $s = 12.2°$.

However, of more general concern is the prediction of melting points of members of diverse data sets. Cramer's [14] $BC(DEF)$ approach, which was quite successful for boiling points, did not perform well on melting points; for a set of 112 chemicals, the standard error was 39.6°, and as similarity to the training set decreased, the standard error increased considerably.

Todeschini et al. [81] used WHIM descriptors to model the melting points of 94 European Union environmental priority chemicals, with $r^2 = 0.834$ and $s = 32.8°$. Katritzky et al. [143] correlated melting points of 443 substituted benzenes

with their CODESSA descriptors, with $r^2 = 0.837$ and $s = 30.2°$, whereas Pogliani [73] used modified connectivities and physicochemical descriptors to model a diverse group of 62 chemicals ($r^2 = 0.856$, $s = 18.0°$).

However, by far the most extensive modeling studies of the melting points of diverse compounds have been those of Yalkowsky and coworkers. Starting from the thermodynamic relationship $T_m = \Delta H_m / \Delta S_m$ (i.e., the ratio of the enthalpy and entropy of melting), Tsakanikas and Yalkowsky [127] incorporated terms to account for conformational flexibility and rotational symmetry, and showed that these, together with group (enthalpic) contributions (first used to predict melting points by Wąchalewski [153]), modeled the melting points of 102 alkanes well ($r^2 = 0.945$, $s = 19.5°$). Abramowitz and Yalkowsky [125] introduced an eccentricity term to account for the entropy of expansion on melting, and showed for a series of 85 rigid, non–hydrogen-bonded compounds that entropic terms, together with boiling point and a correction for ortho-substitution, modeled melting point well ($r^2 = 0.880$, $s = 22.8°$). Abramowitz and Yalkowsky [126] also correlated the melting points of 58 PCBs with entropic terms, the number of chlorine atoms, and terms for ortho-substitution ($r^2 = 0.828$, $s = 22.1°$). Simamora and coworkers, who used group contributions to account for the enthalpy of melting and rotational symmetry to account for the entropy of melting, predicted the melting points of 520 rigid non–hydrogen-bonding organic compounds with $s = 39.2°$ [146] and of 1,181 rigid aromatic compounds (including many with hydrogen bonding capability) with $r^2 = 0.991$ and $s = 36.6°$ [144]. This was then extended [19] to 1,690 aromatic compounds with $r^2 = 0.991$ and $s = 37.5°$, to 596 non–hydrogen-bonding aliphatic compounds [27] with $r^2 = 0.977$ and $s = 34.3°$, and to 1,040 aliphatic compounds [145] with $s = 34.4°$.

Two other laboratories also have developed group contribution methods for melting point prediction. Constantinou and Gani [21] used two levels of group contributions (simple and polyfunctional/structural) to model the melting points of 312 diverse compounds with a standard error of 18.3°, and an average absolute error of 14.0° compared with one of 22.6° when using the method of Joback and Reid [20]. Marrero and Gani [22] recently refined and extended the method and were able to predict the melting points of 1,103 compounds with $s = 25.3°$. Tu [132] was able to predict the melting points of 307 hydrocarbons with an average absolute error of 7.4%; Tu and Wu [149] extended this approach to a set of 1,310 diverse compounds with an average absolute error of 8.2%.

Thus, the work so far on melting point prediction can be seen to have four main thrusts: group contributions, physicochemical and structural properties, topological indices, and entropic contributions. Each has proved its worth, but yet the standard error of prediction remains too high, so that, for example, whether or not a compound will be liquid at 298 K cannot be predicted accurately. It is suggested that a combination of the four approaches might well yield better predictions, with lower standard errors (perhaps even as low as 10°).

Three commercially available programs were found that are able to predict melting point. Experimental melting points could be found for only 96 of the 100 compounds in the test set; they were taken from the *Handbook of Chemistry and Physics* [154] and the ChemFinder Web site [http://chemfinder.cambridgesoft.com]. The experimental melting points ranged from -187.7 to $+27.5°C$. Because these were strongly biased to the low side, it was initially considered that it might be better to use a different set of test compounds with higher melting points. However, when such a test set of 50 compounds was constructed, the predictions were no better; therefore, for reasons of consistency, the decision was made to use the same 100 compounds to test melting point predictions as were used for boiling point and vapor pressure predictions. The experimental and predicted values of boiling point, vapor pressure and melting point for the test set of 100 compounds are given in Table 2.

MPBPVP from Syracuse Research Corporation. This program calculates melting point by two methods, that of Joback and Reid [20] and that of Gold and Ogle [26], and takes their mean. The average absolute prediction error was 26.3°; 23 compounds had errors between 20° and 40°, 18 compounds had errors between 40° and 60°, and 9 compounds had errors $> 60°$ (the worst was benzene with an error of 83.4°).

ChemOffice from CambridgeSoft. This program uses the method of Joback and Reid [20]. It did not calculate a melting point for pyrrole. The average absolute prediction error was 27.0°; 27 compounds had errors between 20° and 40°, 9 compounds had errors between 40° and 60°, and 11 compounds had errors $> 60°$ (the worst was benzene with an error of 107.8°).

ProPred from the Technical University of Denmark. This program did not calculate melting points for formaldehyde and chlorodifluoromethane. The average absolute prediction error was 25.8°; 24 compounds had errors between 20° and 40°, 11 compounds had errors between 40° and 60°, and 10 compounds had errors $> 60°$ (the worst was 2,2,3-trimethylbutane with an error of 96.4°).

REFERENCES

1. Müller M, Klein W. 1993. Fugacity calculations using estimated physicochemical properties. *SAR/QSAR Environ Res* 1:245–262.
2. Yalkowsky SH, Banerjee S. 1992. *Aqueous Solubility: Methods of Estimation for Organic Compounds.* Marcel Dekker, New York, NY, USA.
3. Hansch C, Maloney PP, Fujita T, Muir RM. 1962. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* 194:178–180.
4. Mills EJ. 1884. On melting point and boiling point as related to composition. *Phil Mag* (Ser 5) 17:173–187.
5. Zhao L, Li P, Yalkowsky SH. 1999. Predicting the entropy of boiling for organic compounds. *J Chem Inf Comput Sci* 39:1112–1116.
6. Rücker G, Rücker C. 1999. On topological indices, boiling points, and cycloalkanes. *J Chem Inf Comput Sci* 39:788–802.
7. Horvath AL. 1992. *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds.* Elsevier, Amsterdam, The Netherlands.
8. Nendza M. 1998. *Structure–Activity Relationships in Environmental Sciences.* Chapman & Hall, London, UK.
9. Lyman WJ. 2000. Boiling point. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals.* Lewis, Boca Raton, FL, USA, pp 29–51.
10. Meissner HP. 1949. Critical constants from parachor and molar refraction. *Chem Eng Prog* 45:149–153.
11. Rechsteiner CE. 1990. Boiling point. In Lyman WJ, Reehl WF, Rosenblatt DH, eds, *Handbook of Property Estimation Methods.* Section 12. American Chemical Society, Washington, DC, pp 1–55.
12. Banks WH. 1939. Considerations of a vapour pressure–temper-

ature equation, and their relation to Burnop's boiling point function. *J Chem Soc* 292–295.

13. Burnop VCE. 1938. Boiling point and chemical constitution. Part I. An additive function of molecular weight and boiling point. *J Chem Soc* 826–829.

14. Cramer RD. 1980. *BC(DEF)* parameters. 2. An empirical structure-based scheme for the prediction of some physical properties. *J Am Chem Soc* 102:1849–1859.

15. Lai WY, Chen DH, Maddox RN. 1987. Application of a nonlinear group-contribution model to the prediction of physical constants. 1. Predicting normal boiling points with molecular structure. *Ind Eng Chem Res* 26:1072–1079.

16. Stein SE, Brown RL. 1994. Estimation of normal boiling points from group contributions. *J Chem Inf Comput Sci* 34:581–587.

17. Yalkowsky SH, Krzyzaniak J, Myrdal PB. 1994. Relationships between melting point and boiling point of organic compounds. *Ind Eng Chem Res* 33:1872–1877.

18. Walters AE, Myrdal PB, Yalkowsky SH. 1995. A method for estimating the boiling points of organic compounds from their melting points. *Chemosphere* 31:3001–3008.

19. Simamora P, Yalkowsky SH. 1994. Group contribution methods for predicting the melting points and boiling points of aromatic compounds. *Ind Eng Chem Res* 33:1405–1409.

20. Joback KG, Reid RC. 1987. Estimation of pure-component properties from group contributions. *Chem Eng Commun* 57:233–243.

21. Constantinou L, Gani R. 1994. New group contribution method for estimating properties of pure compounds. *AICHE J* 40:1697–1710.

22. Marrero J, Gani R. 2001. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibr* 183–184:183–208.

23. Katritzky AR, Mu L, Lobanov VS, Karelson M. 1996. 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* 100:10400–10407.

24. Katritzky AR, Lobanov VS, Karelson M. 1998. Normal boiling points for organic compounds: Correlation and prediction by a quantitative structure–property relationship. *J Chem Inf Comput Sci* 38:28–41.

25. Katritzky AR, Maran U, Lobanov VS, Karelson M. 2000. Structurally diverse quantitative structure–property correlations of technologically relevant physical properties. *J Chem Inf Comput Sci* 40:1–18.

26. Gold PI, Ogle GJ. 1969. Estimating thermophysical properties of liquids. Part 4—Boiling, freezing and triple-point temperatures. *Chem Eng* 76:119–122.

27. Krzyzaniak JF, Myrdal PB, Simamora P, Yalkowsky SH. 1995. Boiling point and melting point prediction for aliphatic non–hydrogen-bonding compounds. *Ind Eng Chem Res* 34:2530–2535.

28. Estrada E. 1996. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. *J Chem Inf Comput Sci* 36:844–849.

29. Estrada E, Guevara N, Gutman I. 1998. Extension of edge connectivity index. Relationships to line graph indices and QSPR applications. *J Chem Inf Comput Sci* 38:428–431.

30. Estrada E, Guevara N, Gutman I, Rodriguez L. 1998. Molecular connectivity indices of iterated line graphs. A new source of descriptors for QSPR and QSAR studies. *SAR/QSAR Environ Res* 9:229–240.

31. Basak SC, Grunwald GD, Niemi GJ. 1997. Use of graph-theoretic and geometric descriptors in structure–activity relationships. In Balaban AT, ed, *From Chemical Topology to Three-Dimensional Geometry*. Plenum, New York, NY, USA, pp 73–116.

32. Liu S, Cao C, Li Z. 1998. Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance-edge (MDE) vector, λ. *J Chem Inf Comput Sci* 38:387–394.

33. Ivanciuc O, Ivanciuc T, Cabrol-Bass D, Balaban AT. 2000. Evaluation in quantitative structure–property relationship models of structural descriptors derived from information-theory operators. *J Chem Inf Comput Sci* 40:631–643.

34. Baskin II, Palyulin VA, Zefirov NS. 1997. A neural device for searching direct correlations between structures and properties of chemical compounds. *J Chem Inf Comput Sci* 37:715–721.

35. Arupjyoti S, Iragavarapu S. 1998. New electrotopological descriptor for prediction of boiling points of alkanes and aliphatic alcohols through artificial neural network and multiple linear regression analysis. *Comput Chem* 22:515–522.

36. Thanikaivelan P, Subramanian V, Rao JR, Nair BU. 2000. Application of quantum chemical descriptor in quantitative structure–activity and structure property relationship. *Chem Phys Lett* 323:59–70.

37. Lučić B, Amić D, Trinajstić N. 2000. Nonlinear multivariate regression outperforms several concisely designed neural networks on three QSPR data sets. *J Chem Inf Comput Sci* 40:403–413.

38. Bonchev D. 2000. Overall connectivities/topological complexities: A new powerful tool for QSPR/QSAR. *J Chem Inf Comput Sci* 40:934–941.

39. Hosoya H, Gotoh M, Murakami M, Ikeda S. 1999. Topological index and thermodynamic properties. 5. How can we explain the topological dependency of thermodynamic properties of alkanes with the topology of graphs? *J Chem Inf Comput Sci* 39:192–196.

40. Golovanov IB, Tsygankova IG. 2000. Estimation of physicochemical properties from the structure–property relationship: A new approach. *Quant Struct–Act Relat* 19:554–564.

41. Marano JJ, Holder GD. 1997. General equation for correlating the thermophysical properties of *n*-paraffins, *n*-olefins, and other homologous series. 2. Asymptotic behavior correlations for *PVT* properties. *Ind Eng Chem Res* 36:1895–1907.

42. Trinajstić N, Nikolić S, Basak SC, Lukovits I. 2001. Distance indices and their hyper-counterparts: Intercorrelation and use in the structure–property modelling. *SAR/QSAR Environ Res* 12:31–54.

43. Araujo O, Morales DA. 1998. Properties of new orthogonal graph theoretical invariants in structure–property correlations. *J Chem Inf Comput Sci* 38:1031–1037.

44. Basak SC, Nikolić S, Trinajstić N, Amić D, Bešlo D. 2000. QSPR modeling: Graph connectivity indices versus line graph connectivity indices. *J Chem Inf Comput Sci* 40:927–933.

45. Lučić B, Trinajstić N. 1997. New developments in QSPR/QSAR modeling based on topological indices. *SAR/QSAR Environ Res* 7:45–62.

46. Zhang R, Kiu S, Liu M, Hu Z. 1997. Neural networks—molecular descriptors approach to the prediction of properties of alkenes. *Comput Chem* 21:335–342.

47. Todeschini R, Gramatica P. 1997. The WHIM theory: New 3D molecular descriptors for QSAR in environmental modelling. *SAR/QSAR Environ Res* 7:89–115.

48. Espinosa G, Yaffe D, Cohen Y, Arenas A, Giralt F. 2000. Neural network based quantitative structure property relations (QSPRs) for predicting boiling points of aliphatic hydrocarbons. *J Chem Inf Comput Sci* 40:859–879.

49. Cholakov GS, Wakeham WA, Stateva RP. 1999. Estimation of normal boiling points of hydrocarbons from descriptors of molecular structure. *Fluid Phase Equilibr* 163:21–42.

50. Cao C, Liu S, Li Z. 1999. On molecular polarizability: 2. Relationship to the boiling point of alkanes and alcohols. *J Chem Inf Comput Sci* 39:1105–1111.

51. Hall LH, Story CT. 1997. Boiling point of a set of alkanes, alcohols and chloroalkanes: QSAR with atom type electrotopological state indices using artificial neural networks. *SAR/QSAR Environ Res* 6:139–161.

52. Lin Z, Xu J, Zheng X, Li Z. 2000. Study on quantitative structure-property relationship of chain hydrocarbons, aldehydes and alkanones by molecular distance-edge vector. *Acta Phys–Chem Sin* 16:153–161.

53. Estrada E. 1997. Spectral moments of the edge adjacency matrix in molecular graphs. 2. Molecules containing heteroatoms and QSAR applications. *J Chem Inf Comput Sci* 37:320–328.

54. Ivanciuc O, Ivanciuc T, Balaban AT. 1998. Quantitative structure–property relationship study of normal boiling points for halogen-/ oxygen-/ sulfur-containing organic compounds using the CODESSA program. *Tetrahedron* 54:9129–9142.

55. Carlton TS. 1998. Correlation of boiling points with molecular structure for chlorofluoroethanes. *J Chem Inf Comput Sci* 38:158–164.

56. Randić M, Basak SC. 1999. Optimal molecular descriptors based on weighted path numbers. *J Chem Inf Comput Sci* 39:261–266.

57. Randić M, Basak SC. 2001. A new descriptor for structure–property and structure–activity correlations. *J Chem Inf Comput Sci* 41:650–656.

58. Liu S, Liu H, Xia Z, Cao C, Li Z. 1999. Molecular distance-edge vector (μ): An extension from alkanes to alcohols. *J Chem Inf Comput Sci* 39:951–957.

59. Sharma V, Goswami R, Madan AK. 1997. A novel highly discriminating topological descriptor for structure–property and structure–activity studies. *J Chem Inf Comput Sci* 37: 273–282.

60. Kier LB, Hall LH. 1999. *Molecular Structure Description—The Electrotopological State.* Academic, San Diego, CA, USA.

61. Gironés X, Amat L, Robert D, Carbó-Dorca R. 2000. Use of electron–electron repulsion energy as a molecular descriptor in QSAR and QSPR studies. *J Comput-Aided Mol Des* 14:477–485.

62. Balaban AT, Basak SC, Mills D. 1999. Normal boiling points of 1,ω-alkanedinitriles: The highest increment in a homologous series. *J Chem Inf Comput Sci* 39:769–774.

63. Todeschini R, Gramatica P. 1997. 3D-modelling and prediction by WHIM descriptors. Part 6. Application of WHIM descriptors in QSAR studies. *Quant Struct–Act Relat* 16:120–125.

64. Bünz AP, Braun B, Janowsky R. 1998. Application of quantitative structure–performance relationship and neural network models for the prediction of physical properties from molecular structure. *Ind Eng Chem Res* 37:3043–3051.

65. Bünz AP, Braun B, Janowsky R. 1999. Quantitative structure–property relationships and neural networks: Correlation and prediction of physical properties of pure components and mixtures from molecular structure. *Fluid Phase Equilibr* 158–160:367–374.

66. Chiorboli C, Gramatica P, Piazza R, Pino A, Todeschini R. 1997. 3D-modelling and prediction by WHIM descriptors. Part 7. Physico-chemical properties of haloaromatics: Comparison between WHIM and topological descriptors. *SAR/QSAR Environ Res* 7:133–150.

67. Jain N, Yalkowsky SH. 1999. UPPER III: Unified physical property estimation relationships. Application to non-hydrogen bonding aromatic compounds. *J Pharm Sci* 88:852–860.

68. Randić M, Basak SC. 2000. Construction of high-quality structure–property–activity regressions: The boiling points of sulfides. *J Chem Inf Comput Sci* 40:899–905.

69. Ivanciuc O, Ivanciuc T, Balaban AT. 1998. Design of topological indices. Part 10. Parameters based on electronegativity and covalent radius for the computation of molecular graph descriptors for heteroatom-containing molecules. *J Chem Inf Comput Sci* 38:395–401.

70. Balaban AT, Mills D, Basak SC. 1999. Correlation between structure and normal boiling points of acyclic carbonyl compounds. *J Chem Inf Comput Sci* 39:758–764.

71. Collette TW. 1997. Predicting environmental fate parameters with infrared spectroscopy. *Trends Anal Chem* 16:24–36.

72. Tetteh J, Suzuki T, Metcalfe E, Howells S. 1999. Quantitative structure–property relationships for the estimation of boiling point and flash point using a radial basis function neural network. *J Chem Inf Comput Sci* 39:491–507.

73. Pogliani L. 1999. Modeling with semiempirical connectivity terms. *J Phys Chem A* 103:1598–1610.

74. Stanton DT. 2000. Development of a quantitative structure–property relationship model for estimating normal boiling points of small multifunctional organic molecules. *J Chem Inf Comput Sci* 40:81–90.

75. Labute P. 2000. A widely applicable set of descriptors. *J Mol Graph Model* 18:464–477.

76. Wessel MD, Jurs PC. 1995. Prediction of normal boiling points for a diverse set of industrially important organic compounds from molecular structure. *J Chem Inf Comput Sci* 35:841–850.

77. Goll ES, Jurs PC. 1999. Prediction of the normal boiling points of organic compounds from molecular structure with a computational neural network model. *J Chem Inf Comput Sci* 39: 974–983.

78. Basak SC, Gute BD, Grunwald GD. 1996. A comparative study of topological and geometrical parameters in estimating normal boiling point and octanol/water partition coefficient. *J Chem Inf Comput Sci* 36:1054–1060.

79. Basak SC, Mills D. 2001. Use of mathematical structural invariants in the development of QSPR models. *Commun Math Comput Chem* (44):15–30.

80. Hall LH, Story CT. 1996. Boiling point and critical tempereaure of a heterogeneous data set: QSAR with atom type electrotopological state indices using artificial neural networks. *J Chem Inf Comput Sci* 36:1004–1014.

81. Todeschini R, Vighi M, Finizio A, Gramatica P. 1997. 3D-modelling and prediction by WHIM descriptors. Part 8. Toxicity and physico-chemical properties of environmental priority chemicals by 2D-TI and 3D-WHIM descriptors. *SAR/QSAR Environ Res* 7:173–193.

82. Gramatica P, Navas N, Todeschini R. 1999. Classification of organic solvents and modelling of their physico-chemical properties by chemometric methods using different sets of molecular descriptors. *Trends Anal Chem* 18:461–471.

83. Cocchi M, De Benedetti PG, Seeber R, Tassi L, Ulrici A. 1999. Development of quantitative structure–property relationships using calculated descriptors of the physicochemical properties (n_D, ρ, bp, ε, η) of a series of organic solvents. *J Chem Inf Comput Sci* 39:1190–1203.

84. Topliss JG, Costello RJ. 1972. Chance correlations in structure–activity studies using multiple regression analysis. *J Med Chem* 15:1066–1069.

85. Dearden JC. 2003. An introduction to quantitative structure–activity relationships (QSAR). In Sinha DK, Basak SC, Mohanty RK, Basumallick IN, eds, *Some Aspects of Mathematical Chemistry.* Visva-Bharati University, Santiniketan, India (in press).

86. Alarie Y, Nielsen GD, Andonian-Haftvan J, Abraham MH. 1995. Physicochemical properties of unreactive volatile organic chemicals to estimate RD50: Alternatives to animal studies. *Toxicol Appl Pharmacol* 134:92–99.

87. Abraham MH, Andonian-Haftvan J. 1996. An analysis of nasal irritation thresholds using a new solvation equation. *Fundam Appl Toxicol* 31:71–76.

88. Sage ML, Sage GW. 2000. Vapor pressure. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals.* Lewis, Boca Raton, FL, USA, pp 53–65.

89. Lyman WJ. 1985. Estimation of physical properties. In Neely WB, Blau GE, eds, *Environmental Exposure from Chemicals.* CRC, Boca Raton, FL, USA, pp 13–47.

90. Antoine C. 1888. Tensions des vapeurs; nouvelle rélation entre les tensions et les températures. *Compt Rend* 107:681–684.

91. Grain CF. 1982. Vapor pressure. In Lyman WJ, Reehl WF, Rosenblatt DH, eds, *Handbook of Chemical Property Estimation Methods.* American Chemical Society, Washington, DC, pp 14-1–14-20.

92. Yalkowsky SH, Dannenfelser R-M, Myrdal P, Simamora P, Mishra D. 1994. Unified physical property estimation relationships (UPPER). *Chemosphere* 28:1657–1673.

93. Abrams DS, Massaldi HA, Prausnitz JM. 1974. Vapor pressures of liquids as a function of temperature. Two-parameter equation based on kinetic theory of fluids. *Ind Eng Chem Fundam* 13: 259–262.

94. Macknick AB, Winnick J, Prausnitz JM. 1977. Vapor pressures of liquids as a function of temperature. Two-parameter equation based on kinetic theory of fluids. *Ind Eng Chem Fundam* 16: 392.

95. Macknick AB, Prausnitz JM. 1979. Vapor pressures of heavy liquid hydrocarbons by a group-contribution method. *Ind Eng Chem Fundam* 18:348–351.

96. Burkhard LP. 1985. Estimation of vapor pressures for halogenated aromatic hydrocarbons by a group-contribution method. *Ind Eng Chem Fundam* 24:119–120.

97. Burkhard LP, Andren AW, Armstrong DE. 1985. Estimation of vapor pressures for polychlorinated biphenyls: A comparison of eleven predictive methods. *Environ Sci Technol* 19:500–507.

98. Hawker DW. 1989. Vapor pressures and Henry's law constants of polychlorinated biphenyls. *Environ Sci Technol* 23:1250–1253.

99. Rouvray DH, Tatong W. 1989. Novel applications of topological indices. 3. Prediction of the vapor pressure in polychlorinated biphenyls. *Int J Environ Stud* 33:247–257.

100. Rordorf BF. 1997. Prediction of vapor-pressures, boiling points

and enthalpies of fusion for 29 halogenated dibenzo-*p*-dioxins. *Thermochim Acta* 112:117–122.

101. Banerjee S, Howard PH, Lande SS. 1990. General structure–vapor pressure relationships for organics. *Chemosphere* 21:1173–1180.

102. Kamlet MJ, Doherty RM, Carr PW, Mackay D, Abraham MH, Taft RW. 1988. Linear solvation energy relationships. 44. Parameter-estimation rules that allow accurate prediction of octanol-water partition coefficient and other solubility and toxicity properties of polychlorinated-biphenyls and polycyclic aromatic-hydrocarbons. *Environ Sci Technol* 22:503–509.

103. Tiegs D, Gmehling J, Rasmussen P, Fredenslund A. 1987. Vapor–liquid-equilibria by UNIFAC group contribution. 4. Revision and extension. *Ind Eng Chem Res* 26:159–161.

104. Tu C-H. 1994. Group-contribution method for the estimation of vapor pressures. *Fluid Phase Equilibr* 99:105–120.

105. Simmons KA. 1999. A simple structure-based calculator for estimating vapor pressure. *J Agric Food Chem* 47:1711–1716.

106. Yaffe D, Cohen Y. 2001. Neural network based temperature-dependent quantitative structure property relations (QSPRs) for predicting vapor pressure of hydrocarbons. *J Chem Inf Comput Sci* 41:463–477.

107. Goll ES, Jurs PC. 1999. Prediction of vapor pressures of hydrocarbons and halohydrocarbons from molecular structure with a computational neural network model. *J Chem Inf Comput Sci* 39:1081–1089.

108. Kühne R, Ebert R-U, Schüürmann G. 1997. Estimation of vapour pressures for hydrocarbons and halogenated hydrocarbons from chemical structure by a neural network. *Chemosphere* 34:671–686.

109. Gabriel JL, Miller TF, Wolfson MR, Shaffer TH. 1996. Quantitative structure–activity relationships of perfluorinated heterohydrocarbons as potential respiratory media. *ASAIO J* 42:968–973.

110. Niederfellner J, Lenoir D, Matuschek G, Rehfeldt F, Utschick H, Brüggemann R. 1997. Description of vapour pressures of polycyclic aromatic compounds by graph theoretical indices. *Quant Struct–Act Relat* 16:38–48.

111. Andreev NN, Kuznetsov SE, Storozhenko SY. 1994. Prediction of vapor pressure and boiling points of aliphatic compounds. *Mendeleev Commun* 173–174.

112. Andreev NN, Kuznetsov SE. 1996. Prediction of vapor pressure of volatile corrosion inhibitors. *Prot Metals* 32:148–154 (translation of *Zashch Metallov* 32:163–169).

113. Andreev NN. 1998. To the quantitative estimation of the vapor pressure of volatile corrosion inhibitors. *Prot Metals* 34:101–111 (translation of *Zashch Metallov* 34:123–133).

114. Katritzky AR, Wang Y, Sild S, Tamm T, Karelson M. 1998. QSPR studies on vapor pressure, aqueous solubility, and the prediction of water–air partition coefficients. *J Chem Inf Comput Sci* 38:720–725.

115. Liang CK, Gallagher DA. 1998. QSPR prediction of vapor pressure from solely theoretically-derived descriptors. *J Chem Inf Comput Sci* 38:321–324.

116. Chastrette M, Crétin D. 1995. Structure–property relationships—Determination of the vapor pressure of hydrocarbons and oxygenated compounds using multifunctional autocorrelation method (MAM). *SAR/QSAR Environ Res* 3:131–149.

117. McClelland HE, Jurs PC. 2000. Quantitative structure–property relationships for the prediction of vapor pressures of organic compounds from molecular structures. *J Chem Inf Comput Sci* 40:967–975.

118. Beck B, Breindl A, Clark T. 2000. QM/NN QSPR models with error estimation: Vapor pressure and log*P*. *J Chem Inf Comput Sci* 40:1046–1051.

119. Basak SC, Gute BD, Grunwald GD. 1997. Use of topostructural, topochemical, and geometric parameters in the prediction of vapor pressure: A hierarchical QSAR approach. *J Chem Inf Comput Sci* 37:651–655.

120. Basak SC, Mills D. 2001. Quantitative structure–property relationships (QSPRs) for the estimation of vapor pressure: A hierarchical approach using mathematical structural descriptors. *J Chem Inf Comput Sci* 41:692–701.

121. Cash GG. 1999. Prediction of physicochemical properties from Euclidean distance methods based on electrotopological state indices. *Chemosphere* 39:2583–2591.

122. Delle Site A. 1996. The vapor pressure of environmentally sig-nificant organic chemicals: A review of methods and data at ambient temperature. *J Phys Chem Ref Data* 26:157–193.

123. Dearden JC. 1999. The prediction of melting point. In Charton M, Charton I, eds, *Advances in Quantitative Structure Property Relationships,* Vol 2. JAI, Stamford, CT, USA, pp 127–175.

124. Tesconi M, Yalkowsky SH. 2000. Melting point. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals.* Lewis, Boca Raton, FL, USA, pp 3–27.

125. Abramowitz R, Yalkowsky SH. 1990. Melting point, boiling point and symmetry. *Pharm Res* 7:942–947.

126. Abramowitz R, Yalkowsky SH. 1990. Estimation of aqueous solubility and melting point of PCB congeners. *Chemosphere* 21:1221–1229.

127. Tsakanikas PD, Yalkowsky SH. 1988. Estimation of melting point of flexible molecules: Aliphatic hydrocarbons. *Toxicol Environ Chem* 17:19–33.

128. Pogliani L. 1996. Modeling with special descriptors derived from a medium-sized set of connectivity indices. *J Phys Chem* 100:18065–18077.

129. Pogliani L. 1995. Molecular modeling by linear combinations of connectivity indexes. *J Phys Chem* 99:925–937.

130. Pogliani L. 2000. Modeling with pseudoconnectivity descriptors. A useful extension of the intrinsic I-state concept. *J Phys Chem A* 104:9029–9045.

131. Cherqaoui D, Villemin D, Kvasnicka V. 1994. Application of neural-network approach for prediction of some thermochemical properties of alkanes. *Chemom Intell Lab Syst* 24:117–128.

132. Tu C-H. 1994. Calculation for normal freezing points of hydrocarbons. *J Chin Inst Chem Eng* 25:151–154.

133. Todeschini R, Gramatica P, Provenzani R, Marengo E. 1995. Weighted holistic invariant molecular descriptors. Part 2. Theory development and application on modeling physico-chemical properties of polyaromatic hydrocarbons. *Chemom Intell Lab Syst* 27:221–229.

134. Todeschini R, Gramatica P. 1997. 3D-modelling and prediction by WHIM descriptors. Part 5. Theory development and chemical meaning of WHIM descriptors. *Quant Struct–Act Relat* 16:113–119.

135. Bhattacharjee S, Rao AS, Dasgupta P. 1991. A new index for molecular property correlation in halomethanes. *Comput Chem* 15:319–322.

136. Piazza R, Pino A, Marchini S, Passerini L, Chiorboli C, Tosato ML. 1995. Modelling physico-chemical properties of halogenated benzenes: QSAR optimisation through variables selection. *SAR/QSAR Environ Res* 4:59–71.

137. Gramatica P, Navas N, Todeschini R. 1998. 3D-modelling and prediction by WHIM descriptors. Part 9. Chromatographic relative retention time and physico-chemical properties of polychlorinated biphenyls. *Chemom Int Lab Syst* 40:53–63.

138. Dearden JC. 1991. The QSAR prediction of melting point, a property of environmental relevance. *Sci Total Environ* 109/110:59–68.

139. Katritzky AR, Gordeeva EV. 1993. Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J Chem Inf Comput Sci* 33:835–857.

140. Pogliani L. 1997. Modeling biochemicals with leading molecular connectivity terms. *Med Chem Res* 7:380–393.

141. Murugan R, Grendze MP, Toomey JE, Katritzky AR, Karelson M, Lobanov V, Rachwal P. 1994. Predicting physical properties from molecular structure. *Chemtech* 24:17–23.

142. Medić-Šarić M, Nikolić S, Matijević-Sosa J. 1992. A QSPR study of 3-(phthalimidoalkyl)-pyrazolin-5-ones. *Acta Pharm* 42:153–167.

143. Katritzky AR, Maran U, Karelson M, Lobanov VS. 1997. Prediction of melting points for the substituted benzenes: A QSPR approach. *J Chem Inf Comput Sci* 37:913–919.

144. Simamora P, Miller AH, Yalkowsky SH. 1993. Melting point and normal boiling point correlations: Applications to rigid aromatic compounds. *J Chem Inf Comput Sci* 33:437–440.

145. Zhao L, Yalkowsky SH. 1999. A combined group contribution and molecular geometry approach for predicting melting points of aliphatic compounds. *Ind Eng Chem Res* 38:3581–3584.

146. Simamora P, Yalkowsky SH. 1993. Quantitative structure–property relationship in the prediction of melting point and boiling point of rigid non–hydrogen bonding organic molecules. *SAR/QSAR Environ Res* 1:293–300.

147. Charton M, Charton BI. 1993. Melting points of substituted ethenes, benzenes, and naphthalenes. *Abstract,* 27th Mid Atlantic Regional Meeting, American Chemical Society, Hofstra University, Hempstead, NY, June 3, pp 129–130.

148. Charton M, Charton BI. 1994. Quantitative description of structural effects on melting points of substituted alkanes. *J Phys Org Chem* 7:196–206.

149. Tu C-H, Wu Y-S. 1996. Group-contribution estimation of normal freezing points of organic compounds. *J Chin Inst Chem Eng* 27:323–328.

150. Somayajulu GR 1990. The melting points of ultralong paraffins and their homologues. *Int J Thermophys* 11:555–572.

151. Riazi MR, Al-Sahhaf TA. 1995. Physical properties of *n*-alkanes and *n*-alkylhydrocarbons: application to petroleum mixtures. *Ind Eng Chem Res* 34:4145–4148.

152. Dearden JC, Rahman MH. 1988. QSAR approach to the prediction of melting points of substituted anilines. *Math Comput Model* 11:843–846.

153. Wąchalewski T. 1970. A relation between the melting point of substance and its chemical composition and configuration. *Postępy Fiz* 21:403–410.

154. Lide DR, ed 1997. *Handbook of Chemistry and Physics,* 78th ed. CRC, Boca Raton, FL, USA.