4

Use of Graph-Theoretic and Geometrical Molecular Descriptors in Structure-Activity Relationships

SUBHASH C. BASAK, GREGORY D. GRUNWALD, and GERALD J. NIEMI

Ostensibly there is color, ostensibly sweetness, ostensibly bitterness, but actually only atoms and the void.

GALEN

(Nature and the Greeks, Erwin Schrödinger, 1954)

4.1. INTRODUCTION

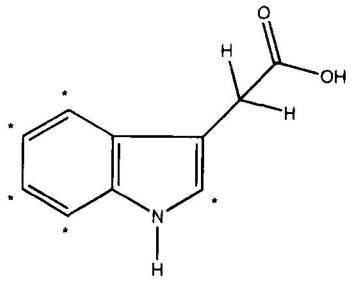
One of the current interests in pharmaceutical drug design,[1–20] chemistry,[21–40] and toxicology[41–53] is the prediction of physicochemical, biomedicinal, and toxicological

SUBHASH C. BASAK, GREGORY D. GRUNWALD, and GERALD J. NIEMI • Center for Water and the Environment, Natural Resources Research Institute, University of Minnesota at Duluth, Duluth, Minnesota 55811.

From Chemical Topology to Three-Dimensional Geometry, edited by Balaban. Plenum Press, New York, 1997

properties of molecules from nonempirical structural parameters which can be calculated directly from their structure. Both in drug design[3,4,31,33,54] and in hazard assessment of chemicals,[31,33,46–53,55] one has to evaluate therapeutic or toxic potential of a large number of compounds, many of which have not even been synthesized. Drug design usually begins with the discovery of a "lead" compound which has the particular therapeutic activity of interest. The lead is altered through molecular modifications and the analogues thus produced are tested until a compound of desirable activity and toxicity profile is found. The combination of possibilities in such a process is almost endless. For example, let us assume the compound in Figure 1 is a lead. The medicinal chemist can carry out numerous manipulations on the lead in terms of substitution. On a very limited scale, if one carries out 50 substitutions in each of the aromatic positions, 10 modifications for esterification, 10 substitutions for the aliphatic carbon and 10 substitutions for the nitrogen, the total number of possible analogues comes to $50^5 \times 10 \times 10 \times 10 = 312.5$ billion structures. This astronomical number is reached by considering only a small fraction of the possible substituents that the medicinal chemist has in his repertoire.[54]

A similar situation exists for the hazard assessment of environmental pollutants. More than 15 million distinct chemical entities have been registered with the Chemical Abstract Service and the list is growing by nearly 775,000 per year. About 1000 of these chemicals enter into societal use every year.[56] Few of these chemicals have experimental properties needed for risk assessment. Table 1 gives a partial list of properties necessary for a reasonable risk assessment of a chemical.[31,33] In the United States, the Toxic Substances Control Act Inventory has about 74,000 entries and the list is growing by nearly 3000 per year. Of the approximately 3000 chemicals



- 50 groups for each aromatic position (*)
- 10 groups for esterification
- 10 groups for aliphatic C
- 10 groups for ring N

Total analogs = 50^5 x 10 x 10 x 10 = 312.5 billion

Figure 1. Probable number of derivatives from a lead via molecular modification.

Table 1. Properties Necessary for Risk Assessment of Chemicals

Physicochemical	Biological
Molar volume	Receptor binding (K_D)
Boiling point	Michaelis constant (K_m)
Melting point	Inhibitor constant (K_i)
Vapor pressure	Biodegradation
Aqueous solubility	Bioconcentration
Dissociation constant (pKa)	Alkylation profile
Partition coefficient	Metabolic profile
Octanol–water (log *P*)	Chronic toxicity
Air–water	Carcinogenicity
Sediment–water	Mutagenicity
Reactivity (electrophile)	Acute toxicity
	LD_{50}
	LC_{50}
	EC_{50}

submitted yearly to the U.S. Environmental Protection Agency for the premanufacture notification process, more than 50% have no experimental data, less than 15% have empirical mutagenicity data, and only about 6% have experimental ecotoxicological and environmental fate data.[55] Also, limited data are available for many of the over 700 chemicals found on the Superfund list of hazardous substances.

In the face of this massive unavailability of experimental data for the vast majority of chemicals, practitioners in drug discovery and hazard assessment have developed the use of nonempirical parameters to estimate molecular properties.[1,3,4,20,31–33] By *nonempirical,* we mean those parameters that can be calculated directly from molecular structure without any other input of experimental data. Topological indexes (TIs), substructural parameters defined on chemical graphs, geometrical (3D or shape) parameters, and quantum-chemical parameters fall in this category.[3,4,21–40,46–55,57–61]

A large number of quantitative structure–activity relationships (QSARs) pertaining to chemistry, pharmacology, and toxicology have used these nonempirical parameters. QSARs are mathematical models that relate molecular structure to their physicochemical, biomedicinal, and toxic properties. Two distinct processes are involved in the derivation of nonempirical parameters for a chemical: (1) defining the model object called "structure" which represents the salient features of the architecture of the chemical species and (2) calculating structural quantifiers from a selected set of critical features of the model object.[31,62] Figure 2 depicts the process of experimental determination of properties vis-à-vis prediction of properties using descriptors.

Figure 2 represents an empirical property as a function $\alpha:C \rightarrow R$ which maps the set C of chemicals into the real line R. A nonempirical QSAR may be regarded as a composition of a description function, $\beta_1:C \rightarrow D$, mapping each chemical structure of C into a space of nonempirical structural descriptors (D) and a prediction function, $\beta_2:D \rightarrow R$, which maps the descriptors into the real line. When $[\alpha(C) - \beta_2\beta_1(C)]$ is within the range of experimental errors, we say that we have a good nonempirical

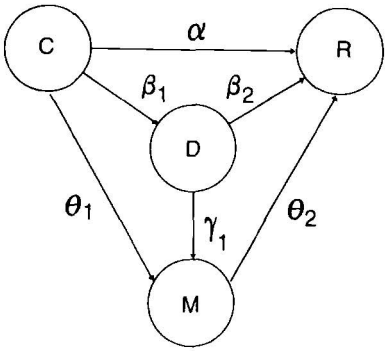


Figure 2. Composition functions for quantitative structure–activity relationship (QSAR) and property– activity relationship (PAR).

predictive model. On the other hand, a property–activity relationship (PAR) is the composition of θ_1:C $\rightarrow$ M, which maps the set C into the molecular property space M, and θ_2:M $\rightarrow$ R, mapping those molecular properties into the real line R. PAR seeks to predict one property (usually a complex property) of a molecule in terms of another (usually simpler or available) property. The latter group of properties may consist either of a number of experimentally determined quantities (e.g., melting point, boiling point, vapor pressure, partition coefficient) or substituent constants or solvatochromic parameters (e.g., steric, electronic, hydrophobic, charge transfer substituent constants, hydrogen bond donor acidity, hydrogen bond acceptor basicity).[54,60] PAR using a calculated property, e.g., calculated partition coefficient (log *P*, octanol-water), may be looked on as a mapping $\theta_2\gamma_1\beta_1$:C $\rightarrow$ R, which is a composition of β_1:C $\rightarrow$ D, γ_1:D $\rightarrow$ M mapping the descriptor space into the molecular property space (e.g., calculation of log *P* from fragments using additivity rule), and θ_2:M $\rightarrow$ R

Graph invariants have been used in a large number of QSARs.[1–53] A graph invariant is a graph-theoretic property that is preserved by isomorphism.[63,64] A graph invariant may be a polynomial, a sequence of numbers, or a single numerical index. Numerical indexes derived from the topological characteristics of molecular graphs are called topological indexes. Molecular structures can be symbolized by graphs where the atomic cores are represented by vertices and covalent chemical bonds are depicted by edges of the graph. Such a graph depicts the connectivity of atoms in a chemical species irrespective of the metric parameters (e.g., equilibrium distance between nuclei, valence angles) associated with the molecular structure. It is in this sense that molecular graphs can be seen as topological, rather than geometrical, representations of molecular structure.[65] TIs are numerical quantifiers of molecular topology and are sensitive to such structural features of molecules as size, shape, symmetry, branching, and cyclicity. Two nonisomorphic graphs may have the same set of graph invariants. In that sense, TIs do not uniquely characterize molecular topology. Yet, it has to be emphasized that TIs quantify many salient aspects of molecular structure. As a result, different graph invariants have been successfully used in characterizing the structural similarity/dissimilarity of molecules,[1–4 28,29,47,49,50,66]

quantifying the degree of molecular branching,[34,35,67] and developing structure-activity relationships in chemistry, biomedical sciences, and environmental toxicology.[5–53,64,67–81]

4.2. TOPOLOGICAL INDEXES AND QSAR

TIs have been used in developing QSAR models for predicting various properties. We give below some examples of successful QSARs using TIs. Definitions of the TIs used in the following equations and throughout this chapter may be found in Table 2.

Table 2. Symbols for Topological Indexes, Geometrical Parameters, and Hydrogen Bonding Parameter and Their Definitions

Index symbol	Definition
I_D^W	Information index for the magnitudes of distances between all possible pairs of vertices of a graph
$\overline{I}_D^W$	Mean information index for the magnitude of distances
${}^v\overline{I}_D^E$	Mean information index for the equality of distances
W	Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph
I^D	Degree complexity
H^V	Graph vertex complexity
H^D	Graph distance complexity
$\overline{IC}$	Information content of the distance matrix partitioned by frequency of occurrences of distance h
O	Order of neighborhood when IC_r reaches its maximum value for the hydrogen-filled graph
I_{ORB}	Information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices
M_1	A Zagreb group parameter = sum of square of degree over all vertices
M_2	A Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices
IC_r	Mean information content or complexity of a graph based on the r^{th} ($r = 0$–6) order neighborhood of vertices in a hydrogen-filled graph
SIC_r	Structural information content for r^{th} ($r = 0$–6) order neighborhood of vertices in a hydrogen-filled graph
CIC_r	Complementary information content for r^{th} ($r = 0$–6) order neighborhood of vertices in a hydrogen-filled graph
TIC_r	Total information content for r^{th} order neighborhood of vertices in a hydrogen-filled graph
${}^h\chi$ *or* ${}^h\chi_p$	Path connectivity index of order $h = 0$–6
${}^h\chi_C$	Cluster connectivity index of order $h = 3$–6
${}^h\chi_{Ch}$	Chain connectivity index of order $h = 3$–6
${}^h\chi_{PC}$	Path–cluster connectivity index of order $h = 4$–6
${}^h\chi^b$	Bonding path connectivity index of order $h = 0$–6
${}^h\chi_C^b$	Bonding cluster connectivity index of order $h = 3$–6
${}^h\chi_{Ch}^b$	Bonding chain connectivity index of order $h = 3$–6
${}^h\chi_{PC}^b$	Bonding path–cluster connectivity index of order $h = 4$–6

(continued)

Table 2. (Continued)

Index symbol	Definition
$^{h}\chi^{v}$	Valence path connectivity index of order $h = 0$–6
$^{h}\chi^{v}_{C}$	Valence cluster connectivity index of order $h = 3$–6
$^{h}\chi^{v}_{Ch}$	Valence chain connectivity index of order $h = 3$–6
$^{h}\chi^{v}_{PC}$	Valence path–cluster connectivity index of order $h = 4$–6
χ_t	Total structure index
Ω-MCI	Orthogonal molecular connectivity indexes
τ	Branchedness indexes
κ	Shape indexes
ϕ	Flexibility indexes
A_3	Half-sum of the cube of the adjacency matrix
p_3	Polarity number: number of third neighbors
N_2	Gordon–Scantlebury index: number of second neighbors
P_h	Number of paths of length $h = 0$–10
J	Balaban's J index based on distance
J^B	Balaban's J index based on multigraph bond orders
J^X	Balaban's J index based on relative electronegativities
J^Y	Balaban's J index based on relative covalent radii
U,V,X,Y	Balaban's information-based indexes on distance sums
AZV	Local vertex invariant based on the adjacency matrix, atomic numbers, and vertex degrees
D	Mean distance topological index for any graph
D_1	Mean distance topological index for acyclic graphs
Z	Hosoya index
HB_1	Hydrogen bonding potential of molecule
ID	Molecular identification numbers
V_W	Volume of molecule
$^{3D}W_H$	3D Wiener number including hydrogens
^{3D}W	3D Wiener number without hydrogens

4.2.1. Physicochemical Properties

4.2.1.1. Boiling Point of Alkanes

Needham *et al.*[21] used TIs to develop a regression equation to predict the normal boiling point (BP) for 74 alkanes:

$$\text{BP} = -9.6 + 38.1(^{1}\chi) - 49.0(1/^{0}\chi) + 5.7(^{4}\chi_{PC}) - 94.5(\chi_t) + 8.4(^{6}\chi_p) \quad (1)$$

$$(N = 74,\ r = 0.999, s = 1.86,\ F = 9030)$$

Subsequently, Basak and Grunwald[78] derived the following equation:

$$\text{BP} = -263 + 237(^{1}\chi) + 18.6(\text{CIC}_2) \quad (2)$$

$$(N = 74,\ r = 0.997, s = 3.83,\ F = 5287)$$

4.2.1.2. Boiling Point of Chlorofluorocarbons (CFCs)

Balaban *et al.*[70] were able to model the boiling points of a large set of CFCs using TIs with the following equation:

$$(3)\quad BP = -73.65 + 33.21({}^{1}\chi^{v} - {}^{0}\chi^{v}) - 64.06({}^{D}\chi^{(0)}) + 94.46({}^{1}\chi) - 20.65(N_{Br}) - 22.18(N_{I}) + 6.36({}^{2}\chi^{v} - {}^{1}\chi^{v})$$
$$(N = 532,\ r = 0.98,\ s = 10.94,\ F = 2953)$$

Using a backpropagation neural network (NN), Balaban *et al.*[25] successfully predicted BP for 276 CFCs. As inputs to the NN, the following parameters were used: *J* index, Wiener index (*W*), number of carbon atoms (N_c), number of chlorine atoms (N_{Cl}), and number of fluorine atoms (N_F). This NN resulted in a correlation(r)=0.992 of observed BP with predicted BP, with a standard error (s) of 8.5°C. The data set used for NN model development consisted of 276 CFCs with, at most, four carbon atoms.

4.2.1.3. Lipophilicity of Diverse Sets of Compounds

Basak *et al.*[27] derived the following equation to predict lipophilicity (log *P*, octanol–water):

$$(4)\quad \log P = 1.76 - 0.50(HB_1) - 5.28(IC_0) - 1.48(CIC_1) + 3.75({}^{0}\chi^{v}) + 0.41(P_6)$$
$$(N = 382,\ r = 0.95,\ s = 0.27,\ F = 1186)$$

where HB_1, is a theoretically calculated hydrogen bonding parameter.

Basak *et al.*[31] developed a refined model for chemicals with HB_1 equal to zero:

$$(5)\quad \log P = -3.13 - 1.64(IC_0) + 2.12({}^{5}\chi_C) - 2.91({}^{6}\chi_{Ch}) + 4.21({}^{0}\chi^{v}) + 1.06({}^{4}\chi^{v}) - 1.02({}^{4}\chi^{v}_{PC})$$
$$(N = 137,\ r = 0.98,\ s = 0.26,\ F = 446)$$

4.2.1.4. Chromatographie Retention Time of Alkanes, Alkylbenzenes

Bonchev and Trinajstić[79] derived the following correlation for alkylbenzenes:

$$(6)\quad RI = 683 + 2.97({}^{v}\overline{I}^{E}_{D}) + 2.71(P_0 - 6)$$
$$N = 28,\ r = 0.99,\ s = 0.58)$$

For alkanes, Kier and Hall[77] found the following relationship:

$$(7)\quad RI = -0.242 + 0.719({}^{1}\chi) + 0.125({}^{3}\chi_p)$$
$$(N = 18,\ r = 0.998,\ s = 0.045,\ F = 1702)$$

4.2.2. Biomedicinal Properties

4.2.2.1. Anesthetic Dose (AD_{50}) of Barbiturates

Basak *et al.*[13] predicted AD_{50} of barbiturates using various TIs:

(8) $$AD_{50} = -49.1 + 200(SIC_1) - 190(SIC_1)^2$$
$$(N = 13, r = 0.76, s = 0.20, F = 6.6)$$

(9) $$AD_{50} = -200 + 153(IC_1) - 28.3(IC_1)^2$$
$$(N = 13, r = 0.74, s = 0.21, F = 6.1)$$

(10) $$AD_{50} = -41.2 + 11.5(^1\chi) - 0.740(^1\chi)^2$$
$$(N = 13, r = 0.72, s = 0.21, F = 5.4)$$

4.2.2.2. Analgesic Potency (A-ED_{50}) of Barbiturates

Basak *et al.*[13] correlated A-ED_{50} of barbiturates using graph-theoretic parameters:

(11) $$\text{A-}ED_{50} = 4700 - 26300(SIC_1) + 36700(SIC_1)^2$$
$$(N = 7, r = 0.97, s = 6.5, F = 29)$$

(12) $$\text{A-}ED_{50} = 5280 - 2800(CIC_1) + 372(CIC_1)^2$$
$$(N = 7, r = 0.96, s = 7.4, F = 27)$$

(13) $$\text{A-}ED_{50} = 2400 - 444(^1\chi) + 20.4(^1\chi)^2$$
$$(N = 7, r = 0.94, s = 9.1, F = 17)$$

4.2.2.3. Enzymatic Acetyl Transfer Reaction

Several TIs have been found to correlate with the enzymatic acetyl transfer reaction,[12] as shown by the following equations:

(14) $$A_\chi = 3.20 - 0.62(^1\chi)$$
$$(N = 9, r = 0.88, s = 0.24, F = 23)$$

(15) $$A_\chi = 2.67 - 0.83(IC_1)$$
$$(N = 9, r = 0.91, s = 0.20, F = 35)$$

(16) $$A_\chi = 3.13 - 4.07(\mathrm{SIC}_1)$$
$$(N = 9, r = 0.92, s = 0.20, F = 36)$$

4.2.2.4. Hill Reaction Inhibitory Potency of Triazinones[6]

(17) $$\mathrm{pI}_{50} = -13.36 + 71.15(\mathrm{SIC}_1) - 63.64(\mathrm{SIC}_1)^2$$
$$(N = 11, r = 0.937, s = 0.316, F = 28.6)$$

4.2.2.5. Complement Inhibition by Benzamidines[80]

(18) $$1/\log_{10} C = -1.125 + 0.487(\mathrm{ID}) + 0.011(O)$$
$$(N = 105, r = 0.941, s = 0.020, F = 391)$$

4.2.2.6. Binding of Barbiturates to Cytochrome P_{450}

Basak[43] used several TIs to correlate the binding of barbiturates to cytochrome P_{450}:

(19) $$K_s = 27.79 - 36.78(\mathrm{IC}_0) + 12.17(\mathrm{IC}_0)^2$$
$$(N = 10, r = 0.99, s = 0.01, F = 156.1)$$

(20) $$K_s = 5.94 - 41.26(\mathrm{SIC}_0) + 71.84(\mathrm{SIC}_0)^2$$
$$(N = 10, r = 0.99, s = 0.01, F = 224.3)$$

(21) $$K_s = 35.74 - 18.45(H^{\mathrm{D}}) + 2.38(H^{\mathrm{D}})^2$$
$$(N = 10, r = 0.98, s = 0.01, F = 94.6)$$

4.2.3. Toxicological Properties

4.2.3.1. Nonspecific Narcotic Activity of Alcohols

Basak and Magnuson[81] correlated the nonspecific narcotic activity (LC_{50}) of alcohols using TIs:

(22) $$\log \mathrm{LC}_{50} = 1.979 - 1.896(\mathrm{CIC}_1)$$
$$(N = 10, r = 0.989, s = 0.323, F = 355.3)$$

4.2.3.2. Nonspecific Toxicity of Esters to *Pimephales promelas*[41]

(23) $$\log \mathrm{LC}_{50} = -0.774 - 0.364(\mathrm{CIC}_1) - 0.774({}^1\chi^{\mathrm{v}})$$
$$(N = 15, r = 0.965, s = 0.194, F = 81.1)$$

$$\log LC_{50} = 1.012 - 0.774(CIC_1) - 0.615(I_D^W) \quad (24)$$
$$(N = 15, r = 0.961, s = 0.204, F = 72.7)$$

4.2.3.3. Mutagenicity of Nitrosamines

Basak *et al.*[42] correlated information- or complexity-based parameters with mutagenic potency of nitrosamines:

$$\ln R = 61.0 - 86.8(IC_0) + 29.2(IC_0)^2 \quad (25)$$
$$(N = 15, r = 0.96, s = 1.17, p < 0.001)$$

$$\ln R = 12.0 - 15.3(IC_1) + 3.84(IC_1)^2 \quad (26)$$
$$(N = 15, r = 0.98, s = 0.86, p < 0.001)$$

4.2.3.4. Mutagenicity of Diverse Structures

Basak *et al.*[46] used six TIs and four substructure (subgraph) indicator variables to develop a linear model to classify a set of 520 diverse chemicals as mutagens or nonmutagens as defined by the Ames mutagenicity test.[82] The data set used in their study consisted of 260 mutagens and 260 nonmutagens. The TIs included three information-based indexes: information content of the graph orbits (I_{ORB}), information content at sixth order (IC_6), and structural information content at zeroth order (SIC_0). A fourth index included number of paths of length 10(P_{10}). The remaining two indexes were connectivity type: third-order bond-corrected cluster connectivity ($^3\chi_C^b$) and third-order valence-corrected chain connectivity ($^3\chi_{Ch}^v$).The four substructure indicators were: (1) nitroso chemicals, (2) halogen-substituted mustard, sulfur mustard, or oxygen mustard, (3) organic sulfate or sulfonate, and (4) a biphenyl amine, benzidine, or 4,4′-methylene dianiline derivative.

Using these parameters, a 74.8% overall correct classification rate was achieved. Jackknifed classification tests showed a 74.6% overall correct classification rate.

4.2.3.5. Toxicity of Monoketones

Basak *et al.*[83] derived the following correlations between TIs and the toxicity (LD_{50}) of monoketones:

$$LD_{50}(\text{control}) = 620.0 - 448.0(CIC_1) + 83.5(CIC_1)^2 \quad (27)$$
$$(N = 13, r = 0.95, s = 9.62, F = 48.9)$$

$$LD_{50}(CCl_4) = 407.0 - 235.0(CIC_0) + 35.1(CIC_0)^2 \quad (28)$$
$$(N = 13, r = 0.97, s = 4.76, F = 74.0)$$

4.2.3.6. Inhibition of *p*-Hydroxylation of Aniline by Alcohols[84]

(29) $$plC_{50} = -13.85 + 25.17(IC_0) - 27.89(SIC_1) - 1.87(CIC_2)$$
$$(N = 20, r = 0.96, F = 62.7)$$

Table 3 gives more exhaustive information about the list of properties of different chemical classes that have been successfully correlated using TIs.

Table 3. Summary of QSARs Using Topological Parameters

Property	Chemical class	Variables[b]	Method	Citation	Ref. No.
BP	Aliphatic alcohols	MCI,ID,J, κ, Elec.	LR	Smeeks and Jurs	85
BP	Alkanes	LOVI/LOIS	NLR	Filip *et al.*	73
BP	Alkanes	MCI	LR	Needham *et al.*	21
BP	Alkanes	^1X	LR	Randić	35
BP	Haloalkanes	W, J, N_{Cl}, N_{Br}, N_F, N_I	NN	Balaban *et al.*	25
BP	Haloalkanes	MCI, N_X	LR	Balaban *et al.*	70
BP	Haloalkanes	N, MCI, κ, ϕ, J	LR	Balaban *et al.*	70
BP	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
BP	Paraffins	Platt's No.	LR	Platt	88
BP	Paraffins	W	LR	Wiener	87
CD	α-Amino acids	MCI	LR	Pogliani	76
CNDO/2 charge	Alkanes	MCI	LR	Hall and Kier	89
Cavity SA	Alcohols	LOVI/LOIS	NLR	Filip *et al.*	73
d	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
d	Alkanes	LOVI/LOIS	NLR	Filip *et al.*	73
d_4^{20}	Infinite linear polymers	W	LR,NLR	Mekenyan *et al.*	69
d_4^{20}	Organophosphorus	MCI	LR	Pogliani	90
d_C	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
ΔG_f	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
ΔH_{vap}	Alkanes	LOVI/LOIS	NLR	Filip *et al.*	73
ΔS	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
Diverse profile	Diverse	TI	CCA	Boecklen and Niemi	91
E_s	Diverse	LOVI/Sub.-TI	LR	Balaban and Catana	92
E_s	Hydrocarbons	$^1\chi^V$	LR	Gupta and Singh	93
ΔH_F	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
ΔH_F	Paraffins	Platt's No.	LR	Platt	88

(continued)

Table 3. (Continued)

Property	Chemical class	Variables[b]	Method	Citation	Ref. No.
MON	Alkanes	J, D, D_1	LR	Balaban	23
MON	Alkanes	LOVI/LOIS	NLR	Filip *et al.*	73
MON	Alkanes	MCI	LR	Pogliani	90
MP	Alkanes	MCI	LR	Pogliani	90
MP	Caffeine homologues	MCI	LR	Pogliani	90
MP	Infinite linear polymers	W	LR,NLR	Mekenyan *et al.*	69
MR	Alkylbenzenes	Ω-MCI	LR	Randic	94
MR	Alkylgermanes	1st-order MCI	LR	Kupchik	75
MR	Heptanes	Ω-MCI	LR	Randić	94
MR	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
MR	Organophosphorus	MCI	LR	Pogliani	90
MR	Paraffins	Platt's No.	LR	Platt	88
MR	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
MV	Paraffins	Platt's No.	LR	Platt	88
MW	α-Amino acids	MCI	LR	Pogliani	76
n_D	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
n_D^{20}	Organophosphorus	MCI	LR	Pogliani	90
P_C	Alkanes	$J, X, Y, V, U, {}^1\chi, AZV$	LR	Balaban and Feroiu	74
P_C	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
R_1	α-Amino acids	MCI	LR	Pogliani	76
RI	Alkanes	MCI	LR	Kier and Hall	77
RI	Alkylbenzenes	I^D, P_0	LR	Bonchev and Trinajstić	79
RI	Diverse drugs	MCI, P_n, κ, Elec.	LR	Rohrbaugh and Jurs	95
RI	Organophosphorus	MCI	LR	Pogliani	90
RON	Alkanes	τ	LR	Pal *et al.*	96
S	α-Amino acids	MCI	LR	Pogliani	76
S	Caffeine homologues	MCI	LR	Pogliani	90
T_C	Alkanes	$J, X, Y, V, U, {}^1\chi, AZV$	LR	Balaban and Feroiu	74
T_C	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86
Ultrasonic sound	Alkanes, alcohols	W, J, MCI, ID	LR	Rouvray and Tatong	98
VP	α-Amino acids	MCI	LR	Pogliani	76
VP	Polychlorinated biphenyls	W, J, MCI, N_{Cl}	LR	Rouvray and Tatong	97
V_C	Alkanes	$J, X, Y, V, U, {}^1\chi, AZV$	LR	Balaban and Feroiu	74
V_C	Nonanes–dodecanes	Z, W, p_3, N_2, A_3	LR	Gao and Hosoya	86

(continued)

Table 3. (Continued)

Property	Chemical class	Variables[b]	Method	Citation	Ref. No.
$\alpha_D 22$	Infinite linear polymers	W	LR,NLR	Mekenyan *et al.*	69
log P	Diverse	TI, HB_1	LR	Basak *et al.*	27
log P	Diverse	TI	LR	Niemi *et al.*	45
log P	Diverse	MCI	LR,NLR, PCR	Niemi *et al.*	99
log P	Diverse, $HB_1 = 0$	TI	LR	Basak *et al.*	31
pI	α-Amino acids	MCI	LR	Pogliani	76
Biomedicinal	Bioactive	Inf.	LR	Ray *et al.*	9
Pharmacological	Bioactive agents	Inf.	LR	Basak *et al.*	5
1/log$C$	Benzamidines	TI	LR	Basak *et al.*	80
A-ED_{50}	Barbiturates	MCI, Inf.	LR	Basak *et al.*	13
AD_{50}	Barbiturates	MCI, Inf.	LR	Basak *et al.*	13
AD_{50}	Barbiturates	MCI, Inf., W	NLR	Basak *et al.*	16
A_X	Anilines	MCI, Inf.	LR	Basak *et al.*	12
Antihistaminic	2-(Piperidin-4-ylamino)-1H-benzimidazoles	Sub.-W	LR	Lukovits	100
BOD	Diverse	MCI	Clustering, DA	Niemi *et al.*	99
BOD	Diverse	MCI, log$P$	Clustering, DA	Niemi *et al.*	101
Biodegradation	Diverse	MCI, κ, Sub.	DA	Gombar and Enslein	102
Carcinogenicity	Diverse	σ, κ, MCI, Sub.	LR,DA	Blake *et al.*	103
Cytostatic activity	1 H-Isoindolediones	Sub.-W	LR	Lukovits	100
Estrogen binding	2-Phenylindoles	Sub.-W	LR	Lukovits	100
K_S	Barbiturates	TI	LR	Basak	43
LC_{50}	Alcohols	CIC	LR	Basak and Magnuson	81
LC_{50}	Esters	W, $^1\chi$, $^1\chi^V$, Inf.	LR	Basak *et al.*	41
LD_{50}	Monoketones	TIC_0, TIC_1, CIC_0, CIC_1	NLR	Basak *et al.*	83
Mutagenicity	Diverse	σ, κ, MCI, Sub.	LR,DA	Blake *et al.*	103
Taste	Sulfamates	Wt-paths	SIMCA	Okuyama *et al.*	104
Therapeutic type	Therapeutics	Wt-paths	Clustering	Randić	40
ln$R$	Nitrosamines	IC_0, IC_1	NLR	Basak *et al.*	42
pIC_{50}	N-alkylnorketobemidones/triazinones	Inf.	LR	Ray *et al.*	6
pIC_{50}	Alcohols	Inf.	LR	Magnuson *et al.*	84

[a]Property: BP = boiling point; CD = crystal density; SA = surface area; d = liquid state density; d_4^{20} = density; d_C = critical density; ΔG_f = free energy of formation; ΔH_{vap} = vaporization enthalpy; ΔS = entropy; E_S = Taft's steric parameter; ΔH_F = heat of formation; MON = motor octane number; MP = melting point; MR = molar refractivity; MV = molar volume; n_D = refractive index; n_D^{20} = refractivity index; P_C = critical pressure; R_l = relaxation rate; RI = retention index; RON = research octane number; S = solubility; T_C = critical temperature; VP = vapor pressure; V_C = critical volume; α_D^{22} = specific rotation; log$P$ = logarithm of the octanol–water partition coefficient; pI = isoelectric points; C = molar concentration of inhibitor required for 50% inhibition of complement; A-ED_{50} = analgesic effective dose; AD_{50} = anesthetic dose; A_χ = enzymatic acetyl transfer reaction rate; BOD = biological oxygen demand; K_S = binding constant; LC_{50} = lethal concentration; LD_{50} = lethal dose; ln$R$ = natural logarithm of the number of revertants per nanomole; pIC_{50} = negative logarithm of the inhibition concentration;

[b]Variables: MCI = molecular connectivity indexes; Inf. = information indexes; LOVI = local vertex invariant; LOIS = local invariant set; Elec. = electronic variables; TI = diverse set of topological indexes; Sub. = substructure.

4.3. TOPOLOGICAL APPROACHES TO MOLECULAR SIMILARITY

One important application of TIs and substructural parameters has been in the quantification of molecular similarity. In practical drug design and risk assessment, good-quality QSARs of specific classes of chemicals, if available, are the best option. However, class-specific QSARs are often not available. In such cases, one selects analogues of the chemical of interest (lead or toxicant), and uses the property of

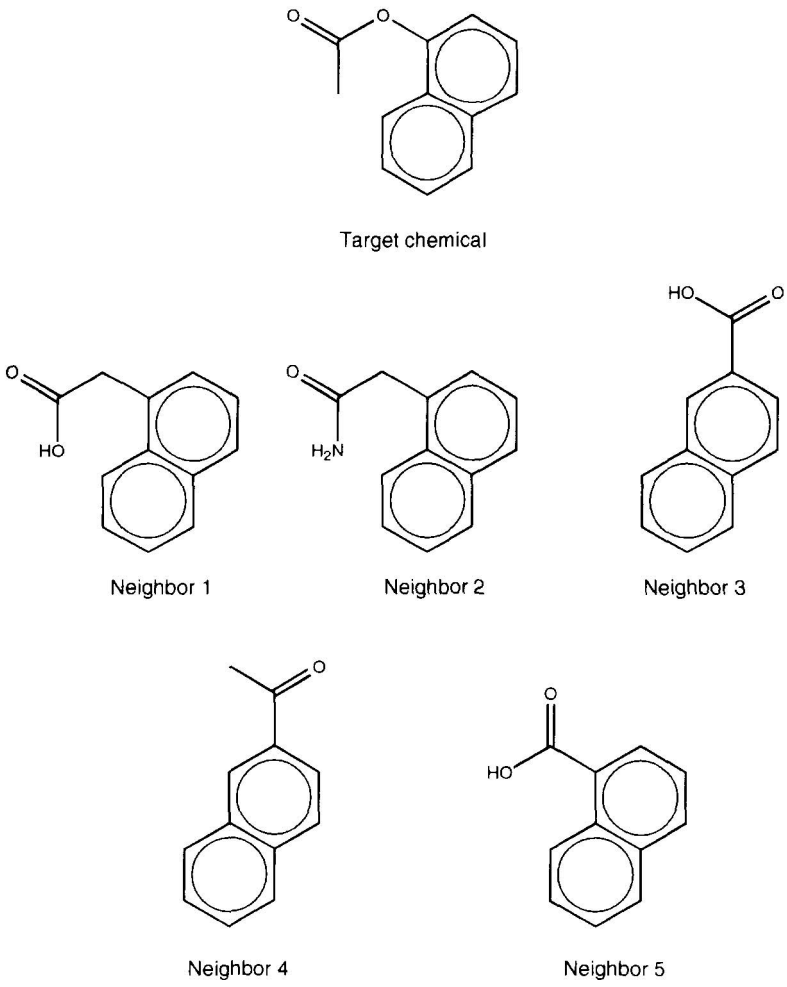


Figure 3. Target chemical and five selected analogues using ED method from the set of 3692 chemicals.

selected analogues for the estimation of the biomedicinal/toxic potential of the chemical.

4.3.1. Quantification of Similarity Using Path Numbers

Path numbers P_h ($h = 1, 2, \ldots$) and weighted paths have been used by Randić and co-workers in determining partial orderings relating dopamine agonist properties for 2-aminotetralins,[105] physicochemical properties of decanes,[106] therapeutic potential of diverse compounds,[40] and antitumor activity of phenyldialkyltriazines.[107] Randić[66] has also reviewed the use of path numbers and weighted paths as they are applied in molecular similarity approaches to property optimization. The results show that the ordering of molecules by path numbers reflects the pattern of activity reasonably well.

4.3.2. Quantification of Similarity Using Topological Indexes

Basak *et al.*[2] used TIs to compute intermolecular similarity of chemicals. Ninety TIs were calculated for a set of 3692 chemicals with diverse structures. Principal component analysis (PCA) was used to reduce the 90-dimensional space to a 10-dimensional subspace which explained 93% of the variance. In the 10-dimensional PC space, the intermolecular similarity of chemicals were quantified in terms of their Euclidean distance (ED). Ten chemicals were then chosen at random from the set of 3692 structures and their analogues were selected using the Euclidean distance as the criterion for nearest-neighbor selection. Figure 3 gives one example of a probe chemical and its five chosen neighbors using this method. The results show that the probe and its selected analogues have a reasonable degree of structural similarity.

4.3.3. Quantification of Intermolecular Similarity Using Substructural Parameters

4.3.3.1. Atom Pairs (APs)

Carhart *et al.*[4] developed the AP method of measuring molecular similarity. An AP is defined as a substructure consisting of two nonhydrogen atoms i and j and their interatomic separation:

$$\langle \text{atom descriptor}_i \rangle\text{-}\langle \text{separation} \rangle\text{-}\langle \text{atom descriptor}_j \rangle$$

where $\langle \text{atom descriptor}_i \rangle$ encodes information about the element type, number of nonhydrogen neighbors, and number of π electrons. Interatomic separation of two atoms is the number of atoms traversed in the shortest bond-by-bond path containing both atoms.

For two molecules, M_i and M_j, AP-based similarity is defined as:

$$S_{ij} = 2C/(T_i + T_j) \tag{30}$$

where C is the number of APs common to molecule i and j. T_i and T_j are the total number of APs in chemicals i and j, respectively. The numerator is multiplied by 2 to reflect the presence of shared APs in both molecules.

The Lederle group has used the AP similarity method to compare chemicals in their data base. Basak *et al.*[28,29,46,47,49,50,53,108] have used the AP method in selecting analogues of chemicals in different and diverse data bases. The relative effectiveness of the AP and ED methods in selecting analogues of chemicals in the STARLIST[109] data base containing more than 4000 chemicals are shown in Figure 4.[108]

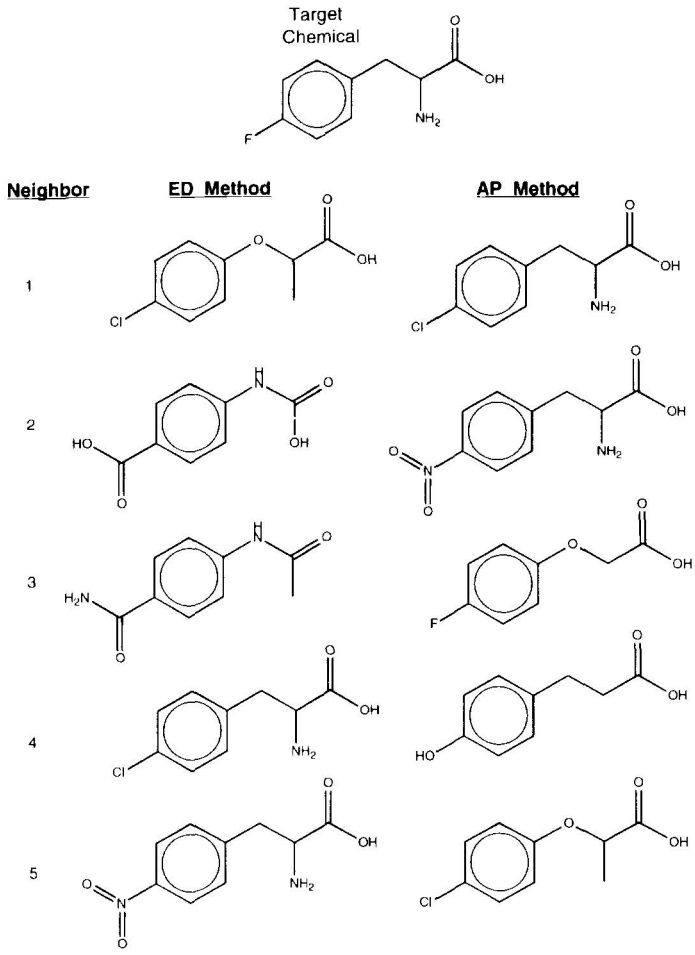


Figure 4. Target chemical and five selected analogues using ED and AP methods from the STARLIST data base of chemicals.

4.3.3.2. Similarity Methods Based on Substructures

Willett and co-workers[110–115] have developed several novel and useful techniques in molecular similarity based on substructural fragments. These approaches are based on the frequency of occurrence of generated fragment descriptors within the molecular graph. Success of these methods has been shown in 2D and 3D matchings of chemical structure, classification of chemical data bases, as well as property estimation.

4.3.4. *K* Nearest-Neighbor (KNN) Method of Estimating Properties

Basak and co-workers also used K (K = 1–10, 15, 20, 25) nearest neighbors of compounds in predicting properties like lipophilicity,[29] boiling point,[28,47,49,116] and mutagenicity[28,47,49,50] of diverse data bases. For a structurally diverse set of 76 compounds, lipophilicity (log$P$, octanol–water) could be reasonably estimated using AP (r = 0.85) and ED (r = 0.85) methods for K= 5.[29]

Four topologically based methods were used by Basak and Grunwald[47] in estimating the boiling point of a set of 139 hydrocarbons and a group of 15 nitrosamines using the nearest neighbor (K = 1).

Basak and Grunwald[50] carried out a comparative study of five molecular similarity techniques, four topologically and one physicochemically based, in estimating the mutagenicity of a set of 73 aromatic and heteroaromatic amines. Of the five methods, two measures of molecular similarity were calculated using topological descriptors, two were derived using physical properties, and the fifth was based on a combination of both topological and physicochemical parameters. The best estimated values were obtained with K = 4–5.

Basak and Grunwald[49] also used topologically based similarity for KNN estimation of the mutagenicity of a set of 95 aromatic amines and the boiling point of a group of over 2900 chemicals with good results.

4.4. GEOMETRICAL/SHAPE PARAMETERS IN SAR

Geometrical parameters, such as molecular shape parameters,[117] sterimol descriptors,[59] volume,[61] bulk parameters,[60,118] and 3D Wiener index,[119] have been developed and used in SARs. Such parameters are derived from the relative distances of atoms in the 3D Euclidean space. We give below some examples of QSARs using 3D descriptors.

4.4.1. van der Waals Volume (V_W)

4.4.1.1. Physicochemical Properties

Bhatnagar *et al.*[120] studied the relationship of boiling point with V_W for several classes of chemicals, including saturated alcohols, primary amines, and alkyl halides:

(31) $$BP_{alcohols} = 5.019 + 127.969(V_W)$$
$$(N = 48, r = 0.964, s = 8.25, F = 605)$$

(32) $$BP_{amines} = -60.175 + 166.419(V_W)$$
$$(N = 21, r = 0.995, s = 5.13, F = 2061)$$

(33) $$BP_{alkyl\ halides} = -108.431 + 226.874(V_W)$$
$$(N = 24, r = 0.896, s = 16.35, F = 90)$$

Correlation of water solubility (molality) with V_W has also been determined for the saturated alcohols[120]:

(34) $$\log S = 6.908 - 8.596(V_W)$$
$$(N = 48, r = 0.974, s = 0.464, F = 860)$$

4.4.1.2. Biomedicinal Properties

Moriguchi and Kanada[61] developed a regression equation modeling the effective concentration (C) of penicillins against *Staphylococcus aureus* in mice:

(35) $$\log (1/C) = 5.911 - 1.692(V_W)$$
$$(N = 18, r = 0.927, s = 0.18)$$

4.4.1.3. Toxicological Properties

For tadpole narcosis of a diverse set of chemicals, the following equation has been developed[61]:

(36) $$\log (1/C) = -2.022 + 2.940(V_W)$$
$$(N = 53, r = 0.969, s = 0.29)$$

Correlation of nonspecific toxicity on the Madison 517 fungus, expressed as $\log(1/C)$ (C is the minimum toxic dose), with V_W was found to be[61]:

(37) $$\log (1/C) = -1.236 + 2.645(V_W)$$
$$(N = 45, r = 0.982, s = 0.19)$$

4.4.2. Comparative Molecular Field Analysis (CoMFA) Approach

In the CoMFA method developed by Cramer *et al.*,[121] a molecule is described using electrostatic, steric, and, sometimes, hydrogen bonding fields calculated at the

intersections of a 3D lattice. The partial least-squares method is used to describe statistical relationships between these fields and biological activity.

4.5. COMPARATIVE STUDY OF TOPOLOGICAL VERSUS GEOMETRICAL DESCRIPTORS IN QSARs

It is clear from the above that both topological and 3D descriptors have been extensively used in QSARs of large sets of molecules. However, no systematic work has been carried out on the relative effectiveness of TIs versus 3D parameters in the prediction of properties using QSAR models. We summarize below the results of our recent studies on the utility of graph-theoretic indexes and geometrical parameters such as 3D Wiener index and volume in estimating: (1) normal boiling point of a set of 140 hydrocarbons, (2) lipophilicity ($\log P$, octanol–water) of a diverse set of 254 molecules, and (3) mutagenic potency ($\ln R$, R being the number of revertants per nanomole in the Ames test) of a set of 95 aromatic and heteroaromatic amines.

4.5.1. Property Data Bases

4.5.1.1. Boiling Point

All normal BP data for the hydrocarbons were found in the literature. The hydrocarbons analyzed include 74 alkanes,[21] 29 alkyl benzenes,[122] and 37 polycyclic aromatic hydrocarbons.[123] Table 4 presents a list of the hydrocarbon compounds with their normal BP (°C).

Table 4. Normal Boiling Point (°C) for 140 Hydrocarbons and Predicted Boiling Point Using Equations (44) and (45)

			Predicted BP	
No.	Chemical name	Obsd. BP	Eq. (44)	Eq. (45)
1	ethane	−88.6	−108.1	−94.7
2	*n*-propane	−42.1	−61.3	−47.7
3	*n*-butane	−0.5	−16.1	−2.3
4	2-methylpropane	−11.7	−17.9	−9.3
5	*n*-pentane	36.1	26.3	36.8
6	2-methylbutane	27.8	21.6	27.6
7	2,2-dimethylpropane	9.5	15.8	22.0
8	*n*-hexane	68.7	64.6	70.5
9	2-methylpentane	60.3	51.8	59.9
10	3-methylpentane	63.3	57.6	64.1
11	2,2-dimethylbutane	49.7	51.9	53.9
12	2,3-dimethylbutane	58.0	61.6	65.0
13	*n*-heptane	98.4	99.0	99.8
14	2-methylhexane	90.0	83.6	88.9

(continued)

Table 4. (Continued)

No.	Chemical name	Obsd. BP	Predicted BP	
			Eq. (44)	Eq. (45)
15	3-methylhexane	91.8	86.5	91.5
16	3-ethylpentane	93.5	91.6	98.3
17	2,2-dimethylpentane	79.2	75.6	80.3
18	2,3-dimethylpentane	89.8	90.9	91.1
19	2,4-dimethylpentane	80.5	83.6	89.3
20	3,3-dimethylpentane	86.1	81.8	83.6
21	2,2,3-trimethylbutane	80.9	88.4	88.4
22	*n*-octane	125.7	129.9	124.7
23	2-methylheptane	117.7	113.4	113.9
24	3-methylheptane	118.9	115.1	116.2
25	4-methylheptane	117.7	114.3	115.8
26	3-ethylhexane	118.5	119.2	123.0
27	2,2-dimethylhexane	106.8	102.7	104.3
28	2,3-dimethylhexane	115.6	113.7	114.5
29	2,4-dimethylhexane	109.4	114.9	112.2
30	2,5-dimethylhexane	109.1	108.8	110.4
31	3,3-dimethylhexane	112.0	105.2	106.4
32	3,4-dimethylhexane	117.7	119.0	117.9
33	2-methyl-3-ethylpentane	115.6	115.0	116.7
34	3-methyl-3-ethylpentane	118.3	111.8	115.6
35	2,2,3-trimethylpentane	109.8	111.7	109.1
36	2,2,4-trimethylpentane	99.2	106.5	105.1
37	2,3,3-trimethylpentane	114.8	115.0	112.3
38	2,3,4-trimethylpentane	113.5	120.6	120.6
39	2,2,3,3-tetramethylbutane	106.5	119.0	112.7
40	*n*-nonane	150.8	158.0	147.3
41	2-methyloctane	143.3	140.8	136.7
42	3-methyloctane	144.2	142.2	138.6
43	4-methyloctane	142.5	140.7	138.4
44	3-ethylheptane	143.0	145.4	145.8
45	4-ethylheptane	141.2	144.7	145.5
46	2,2-dimethylheptane	132.7	128.7	126.2
47	2,3-dimethylheptane	140.5	138.9	136.8
48	2,4-dimethylheptane	133.5	136.7	133.3
49	2,5-dimethylheptane	136.0	137.5	133.8
50	2,6-dimethylheptane	135.2	135.3	132.3
51	3,3-dimethylheptane	137.3	130.4	128.5
52	3,4-dimethylheptane	140.6	141.3	139.1
53	3,5-dimethylheptane	136.0	142.8	138.8
54	4,4-dimethylheptane	135.2	130.2	127.9
55	2-methyl-3-ethylhexane	138.0	139.6	138.9
56	2-methyl-4-ethylhexane	133.8	137.2	136.7
57	3-methyl-3-ethylhexane	140.6	135.9	136.5
58	3-methyl-4-ethylhexane	140.4	145.9	146.1

Table 4. (Continued)

No.	Chemical name	Obsd. BP	Predicted BP	
			Eq. (44)	Eq. (45)
59	2,2,3-trimethylhexane	133.6	131.7	130.0
60	2,2,4-trimethylhexane	126.5	130.2	125.4
61	2,2,5-trimethylhexane	124.1	129.0	124.1
62	2,3,3-trimethylhexane	137.7	134.8	132.9
63	2,3,4-trimethylhexane	139.0	144.9	140.7
64	2,3,5-trimethylhexane	131.3	138.5	137.4
65	2,4,4-trimethylhexane	130.6	133.0	128.0
66	3,3,4-trimethylhexane	140.5	137.5	133.2
67	3,3-diethylpentane	146.2	145.0	144.5
68	2,2-dimethyl-3-ethylpentane	133.8	132.9	130.8
69	2,3-dimethyl-3-ethylpentane	142.0	140.0	136.7
70	2,4-dimethyl-3-ethylpentane	136.7	144.9	139.2
71	2,2,3,3-tetramethylpentane	140.3	141.3	132.8
72	2,2,3,4-tetramethylpentane	133.0	139.3	134.7
73	2,2,4,4-tetramethylpentane	122.3	130.9	128.2
74	2,3,3,4-tetramethylpentane	141.6	145.0	139.2
75	benzene	80.1	99.2	76.0
76	toluene	110.6	121.3	112.2
77	ethylbenzene	136.2	152.6	137.5
78	*o*-xylene	144.4	142.3	146.7
79	*m*-xylene	139.1	137.3	135.1
80	*p*-xylene	138.4	137.3	134.5
81	*n*-propylbenzene	159.2	182.0	163.6
82	1-methyl-2-ethylbenzene	165.2	170.1	169.2
83	1-methyl-3-ethylbenzene	161.3	166.8	158.7
84	1-methyl-4-ethylbenzene	162.0	166.0	157.9
85	1,2,3-trimethylbenzene	176.1	162.7	175.6
86	1,2,4-trimethylbenzene	169.4	161.4	166.5
87	1,3,5-trimethylbenzene	164.7	162.1	167.5
88	*n*-butylbenzene	183.3	209.0	190.4
89	1,2-diethylbenzene	183.4	195.2	192.8
90	1,3-diethylbenzene	181.1	193.0	189.2
91	1,4-diethylbenzene	183.8	194.4	188.1
92	1-methyl-2-*n*-propylbenzene	184.8	194.8	193.5
93	1-methyl-3-*n*-propylbenzene	181.8	191.0	183.5
94	1-methyl-4-*n*-propylbenzene	183.8	190.4	182.5
95	1,2-dimethyl-3-ethylbenzene	193.9	187.4	196.2
96	1,2-dimethyl-4-ethylbenzene	189.8	186.6	187.5
97	1,3-dimethyl-2-ethylbenzene	190.0	187.2	193.5
98	1,3-dimethyl-4-ethylbenzene	188.4	186.4	190.4
99	1,3-dimethyl-5-ethylbenzene	183.8	187.0	188.6
100	1,4-dimethyl-2-ethylbenzene	186.9	187.2	190.7
101	1,2,3,4-tetramethylbenzene	205.0	185.2	202.9
102	1,2,3,5-tetramethylbenzene	198.2	185.1	198.8

(continued)

Table 4. (Continued)

No.	Chemical name	Obsd. BP	Predicted BP	
			Eq. (44)	Eq. (45)
103	1,2,4,5-tetramethylbenzene	196.8	185.7	198.5
104	naphthalene	218.0	228.5	209.9
105	acenaphthalene	270.0	234.4	267.6
106	acenaphthene	279.0	—	272.0
107	fluorene	294.0	299.2	295.7
108	phenanthrene	338.0	346.4	329.8
109	anthracene	340.0	344.8	330.6
110	4*H*-cyclopenta(*def*)phenanthrene	359.0	326.9	349.7
111	fluoranthene	383.0	416.0	378.4
112	pyrene	393.0	392.2	384.3
113	benzo(*a*)fluorene	403.0	386.0	406.4
114	benzo(*b*)fluorene	398.0	390.8	406.4
115	benzo(*c*)fluorene	406.0	386.3	404.5
116	benzo(*ghi*)fluoranthene	422.0	443.4	427.0
117	cyclopenta(*cd*)pyrene	439.0	—	435.7
118	chrysene	431.0	445.8	439.2
119	benz(*a*)anthracene	425.0	440.1	439.7
120	triphenylene	429.0	454.6	433.6
121	naphthacene	440.0	445.2	444.2
122	benzo(*b*)fluoranthene	481.0	503.9	476.0
123	benzo(*j*)fluoranthene	480.0	492.2	474.8
124	benzo(*k*)fluoranthene	481.0	501.1	489.9
125	benzo(*a*)pyrene	496.0	485.9	487.9
126	benzo(*e*)pyrene	493.0	490.8	484.4
127	perylene	497.0	484.0	479.7
128	anthanthrene	547.0	527.7	539.2
129	benzo(*ghi*)perylene	542.0	526.9	529.7
130	indeno(1,2,3-*cd*)fluoranthene	531.0	547.3	541.2
131	indeno(1,2,3-*cd*)pyrene	534.0	540.1	534.5
132	dibenz(*a,c*)anthracene	535.0	535.2	533.7
133	dibenz(*a,h*)anthracene	535.0	528.9	546.2
134	dibenz(*a,j*)anthracene	531.0	529.6	543.6
135	picene	519.0	531.1	545.1
136	coronene	590.0	575.6	591.6
137	dibenzo(*a,e*)pyrene	592.0	574.9	581.6
138	dibenzo(*a,h*)pyrene	596.0	569.8	591.8
139	dibenzo(*a,i*)pyrene	594.0	569.2	591.1
140	dibenzo(*a,l*)pyrene	595.0	573.9	590.5

4.5.1.2. Lipophilicity (log*P*, Octanol–Water)

The 254 chemicals used to model log *P* are presented in Table 5. These chemicals were a subset of 382 chemicals studied by us previously[27] and consist of only those compounds with measured log *P* available in STARLIST,[109] a selected subset of data deemed to be of very high quality by experts in the field. The log *P* values are provided in Table 5.

Table 5. Observed log *P* and Estimated log *P* from Equations (46) and (47) for 254 Diverse Chemicals

			Estimated log *P*	
No.	Chemical name	Obsd. log *P*	Eq. (46)	Eq. (47)
1	butane	2.89	1.61	1.69
2	pentane	3.39	2.08	2.16
3	cyclopentane	3.00	2.61	2.46
4	cyclohexane	3.44	2.27	2.46
5	1-butene	2.40	1.65	1.57
6	1-hexene	3.39	2.64	2.63
7	cyclohexene	2.86	2.44	2.48
8	1-pentyne	1.98	1.66	1.63
9	ethylchloride	1.43	1.05	1.20
10	1-chloropropane	2.04	1.43	1.65
11	1-chlorobutane	2.64	1.81	2.04
12	1-chloroheptane	4.15	3.09	3.32
13	carbon tetrachloride	2.83	2.25	2.49
14	1,2-dichloroethane	1.48	1.60	2.03
15	1,1,1-trichloroethane	2.49	1.88	2.19
16	1,1,2,2-tetrachloroethane	2.39	2.95	3.36
17	trichloroethylene	2.42	2.86	3.16
18	tetrachloroethylene	3.40	3.79	4.04
19	trichlorofluoromethane	2.53	2.66	2.46
20	benzene	2.13	2.21	2.14
21	toluene	2.73	2.59	2.43
22	*o*-xylene	3.12	3.09	2.98
23	*m*-xylene	3.20	3.08	3.00
24	*p*-xylene	3.15	2.89	2.79
25	1,3,5-trimethylbenzene	3.42	3.61	3.58
26	1,2,4-trimethylbenzene	3.78	3.62	3.57
27	1,2,3-trimethylbenzene	3.66	3.60	3.50
28	1,2,3,4-tetramethylbenzene	4.11	3.96	3.84
29	1,2,3,5-tetramethylbenzene	4.17	4.25	4.14
30	1,2,4,5-tetramethylbenzene	4.00	3.98	3.85
31	pentamethylbenzene	4.56	4.70	4.73
32	hexamethylbenzene	5.11	5.15	5.13
33	ethylbenzene	3.15	2.94	2.83

(continued)

Table 5. (Continued)

No.	Chemical name	Obsd. log *P*	Estimated log *P*	
			Eq. (46)	Eq. (47)
34	propylbenzene	3.72	3.37	3.30
35	isopropylbenzene	3.66	3.35	3.30
36	butylbenzene	4.26	3.82	3.82
37	*t*-butylbenzene	4.11	3.76	3.71
38	*p*-cymene	4.10	3.85	3.78
39	fluorobenzene	2.27	2.37	2.00
40	chlorobenzene	2.84	2.37	2.42
41	bromobenzene	2.99	2.37	2.65
42	iodobenzene	3.25	2.37	3.15
43	*o*-dichlorobenzene	3.38	2.97	3.12
44	1,3-dichlorobenzene	3.60	2.97	3.15
45	*p*-dichlorobenzene	3.52	2.89	3.03
46	1,2,3-trichlorobenzene	4.05	3.71	3.92
47	1,2,4-trichlorobenzene	4.02	3.56	3.74
48	1,3,5-trichlorobenzene	4.15	3.66	3.95
49	1,2,3,4-tetrachlorobenzene	4.64	4.34	4.55
50	1,2,3,5-tetrachlorobenzene	4.92	4.32	4.57
51	1,2,4,5-tetrachlorobenzene	4.82	4.31	4.56
52	pentachlorobenzene	5.17	5.15	5.40
53	hexachlorobenzene	5.31	6.05	6.27
54	*o*-dibromobenzene	3.64	2.97	3.52
55	*p*-dibromobenzene	3.79	2.89	3.48
56	*o*-chlorotoluene	3.42	2.87	2.88
57	*m*-chlorotoluene	3.28	2.87	2.90
58	*p*-chlorotoluene	3.33	2.79	2.83
59	ethyl ether	0.89	1.23	1.21
60	dipropyl ether	2.03	2.12	2.15
61	dibutyl ether	3.21	2.91	2.96
62	tetrahydrofuran	0.46	1.82	1.50
63	ethyl vinyl ether	1.04	1.25	1.09
64	anisole	2.11	2.08	1.91
65	*o*-methylanisole	2.74	2.70	2.58
66	*m*-methylanisole	2.66	2.69	2.59
67	*p*-methylanisole	2.81	2.59	2.50
68	4-chloroanisole	2.78	2.36	2.39
69	phenetole	2.51	2.50	2.41
70	phenyl propyl ether	3.18	2.95	2.88
71	formic acid, propyl ester	0.83	1.16	0.92
72	acetic acid, methyl ester	0.18	0.72	0.18
73	acetic acid, ethyl ester	0.73	1.04	0.70
74	propionic acid, ethyl ester	1.21	1.49	1.23
75	acrylic acid, methyl ester	0.80	1.03	0.50
76	methacrylic acid, methyl ester	1.38	1.35	0.91
77	benzoic acid, methyl ester	2.12	2.29	2.08

Table 5. (Continued)

No.	Chemical name	Obsd. log *P*	Estimated log *P*	
			Eq. (46)	Eq. (47)
78	ethyl benzoate	2.64	2.68	2.46
79	*o*-toluic acid, methyl ester	2.75	2.87	2.68
80	acetic acid, benzy lester	1.96	2.70	2.58
81	acetic acid, β-phenylethyl ester	2.30	3.00	2.95
82	phenylacetic acid, methy lester	1.83	2.69	2.53
83	β-phenylpropionic acid, ethyl ester	2.73	3.36	3.32
84	benzyl benzoate	3.97	3.68	3.70
85	acetic acid, phenyl ester	1.49	2.32	2.09
86	*o*-tolylacetate	1.93	2.89	2.72
87	*m*-tolylacetate	2.09	2.81	2.61
88	*p*-tolylacetate	2.11	2.81	2.66
89	2-chlorophenyl acetate	2.18	2.69	2.59
90	3-chlorophenyl acetate	2.32	2.62	2.56
91	2-bromophenyl acetate	2.20	2.69	2.76
92	propionaldehyde	0.59	0.87	0.64
93	butyraldehyde	0.88	1.25	1.16
94	hexaldehyde	1.78	2.16	2.13
95	benzaldehyde	1.48	2.11	1.87
96	acetone	−0.24	0.47	0.17
97	2-butanone	0.29	1.20	0.94
98	2-pentanone	0.91	1.63	1.46
99	2-hexanone	1.38	2.20	2.10
100	2-heptanone	1.98	2.58	2.50
101	cyclohexanone	0.81	1.85	1.87
102	acetophenone	1.58	2.52	2.36
103	*m*-chloroacetophenone	2.51	2.89	2.86
104	*p*-chloroacetophenone	2.32	2.76	2.77
105	*p*-bromoacetophenone	2.43	2.76	2.95
106	*p*-fluoroacetophenone	1.72	2.76	2.42
107	*p*-methylacetophenone	2.10	2.99	2.86
108	propiophenone	2.19	2.92	2.76
109	1-phenyl-2-propanone	1.44	2.95	2.77
110	ethylamine	−0.13	−0.66	−0.51
111	propylamine	0.48	−0.13	0.23
112	butylamine	0.97	0.29	0.70
113	amylamine	1.49	0.81	1.22
114	hexylamine	2.06	1.23	1.64
115	heptylamine	2.57	1.64	2.06
116	diethylamine	0.58	0.72	1.00
117	dipropylamine	1.67	1.64	1.93
118	dibutylamine	2.83	2.43	2.70
119	trimethylamine	0.16	0.11	0.13
120	triethylamine	1.45	1.99	1.95
121	tripropylamine	2.79	3.30	3.25

(continued)

Table 5. (Continued)

No.	Chemical name	Obsd. log *P*	Estimated log *P*	
			Eq. (46)	Eq. (47)
122	aniline	0.90	0.71	0.89
123	*o*-toluidine	1.32	1.31	1.51
124	*m*-toluidine	1.40	1.31	1.54
125	*p*-toluidine	1.39	1.23	1.43
126	*m*-chloroaniline	1.88	1.12	1.46
127	*p*-chloroaniline	1.83	1.04	1.38
128	*m*-bromoaniline	2.10	1.12	1.69
129	*p*-bromoaniline	2.26	1.04	1.62
130	*m*-fluoroaniline	1.30	1.12	1.05
131	*p*-fluoroaniline	1.15	1.04	0.97
132	benzidine	1.34	1.37	1.94
133	α-naphthylamine	2.25	2.20	2.39
134	β-naphthylamine	2.28	2.16	2.31
135	*N,N*-dimethylaniline	2.31	2.45	2.47
136	*N,N*-dimethyl-*p*-toluidine	2.81	2.95	2.96
137	*N,N*-diethylaniline	3.31	3.40	3.44
138	*N,N*-dimethylbenzylamine	1.98	2.90	2.94
139	pyridine	0.65	1.45	1.46
140	3-methylpyridine	1.20	1.69	1.65
141	3-chloropyridine	1.33	1.67	1.81
142	3-bromopyridine	1.60	1.67	2.07
143	4-bromopyridine	1.54	1.67	2.08
144	acetonitrile	–0.34	0.45	0.24
145	propionitrile	0.16	0.93	0.84
146	butyronitrile	0.53	1.27	1.21
147	benzonitrile	1.56	2.17	2.01
148	phenylacetonitrile	1.56	2.54	2.40
149	benzylacetonitrile	1.72	2.96	2.93
150	acrylonitrile	0.25	1.19	0.99
151	nitromethane	–0.35	–0.37	–1.01
152	nitroethane	0.18	0.47	0.10
153	1-nitropropane	0.87	0.75	0.52
154	1-nitrobutane	1.47	1.21	1.03
155	1-nitropentane	2.01	1.57	1.45
156	nitrobenzene	1.85	1.64	1.38
157	*m*-nitrotoluene	2.45	2.13	1.99
158	*p*-nitrotoluene	2.37	2.02	1.87
159	2-chloro-1-nitrobenzene	2.24	2.12	2.02
160	3-chloro-1-nitrobenzene	2.41	2.12	2.07
161	4-chloro-1-nitrobenzene	2.39	2.01	1.98
162	3-bromo-1-nitrobenzene	2.64	2.12	2.30
163	4-bromo-1-nitrobenzene	2.55	2.01	2.20
164	*m*-dinitrobenzene	1.49	1.71	1.43
165	*p*-dinitrobenzene	1.46	1.63	1.38

Table 5. (Continued)

No.	Chemical name	Obsd. log *P*	Estimated log *P*	
			Eq. (46)	Eq. (47)
166	dimethylformamide	−1.01	0.19	−0.03
167	*N,N*-dimethylacetamide	−0.77	0.61	0.44
168	diethylacetamide	0.34	1.74	1.60
169	benzamide	0.64	0.75	0.85
170	dimethylsulfoxide	−1.35	0.07	0.36
171	diethylsulfide	1.95	1.72	2.06
172	methanol	−0.77	−0.13	−0.66
173	ethanol	−0.31	−0.07	−0.30
174	propanol	0.25	0.39	0.43
175	butanol	0.88	0.80	0.97
176	isobutanol	0.76	0.64	0.67
177	pentanol	1.56	1.30	1.50
178	isopentanol	1.42	1.10	1.21
179	hexanol	2.03	1.71	1.89
180	octanol	2.97	2.45	2.70
181	allyl alcohol	0.17	0.38	0.26
182	isopropanol	0.05	−0.00	−0.05
183	*s*-butanol	0.61	0.78	0.82
184	3-pentanol	1.21	1.04	1.07
185	cyclohexanol	1.23	1.49	1.75
186	*t*-butanol	0.35	0.26	0.22
187	2-ethyl-2-propanol	0.89	1.02	1.02
188	benzyl alcohol	1.10	1.59	1.54
189	*m*-methylbenzyl alcohol	1.60	2.20	2.23
190	*p*-methylbenzyl alcohol	1.58	2.10	2.12
191	*m*-chlorobenzyl alcohol	1.94	1.97	2.13
192	*p*-chlorobenzyl alcohol	1.96	1.87	2.02
193	2-phenylethanol	1.36	2.01	2.08
194	3-phenylalcohol	1.88	2.46	2.60
195	cinnamyl alcohol	1.95	2.36	2.38
196	phenol	1.46	1.28	1.19
197	*m*-methylphenol	1.96	1.84	1.83
198	*p*-methylphenol	1.94	1.75	1.73
199	*m*-chlorophenol	2.50	1.70	1.83
200	*p*-chlorophenol	2.39	1.62	1.75
201	*m*-bromophenol	2.63	1.70	2.08
202	*p*-bromophenol	2.59	1.62	1.99
203	*m*-fluorophenol	1.93	1.70	1.40
204	*p*-fluorophenol	1.77	1.62	1.27
205	acetic acid	-0.17	-0.19	-0.74
206	propionic acid	0.33	0.23	-0.13
207	butyric acid	0.79	0.55	0.39
208	valeric acid	1.39	1.07	1.01
209	hexanoic acid	1.92	1.43	1.43

(continued)

Table 5. (Continued)

No.	Chemical name	Obsd. log *P*	Estimated log *P*	
			Eq. (46)	Eq. (47)
210	decanoic acid	4.09	2.69	2.86
211	benzoic acid	1.87	1.47	1.32
212	*m*-toluic acid	2.37	2.01	1.95
213	*p*-toluic acid	2.27	1.88	1.78
214	*m*-chlorobenzoic acid	2.68	1.89	1.93
215	*p*-chlorobenzoic acid	2.65	1.76	1.84
216	*m*-bromobenzoic acid	2.87	1.89	2.13
217	*p*-bromobenzoic acid	2.86	1.76	2.01
218	*m*-fluorobenzoic acid	2.15	1.89	1.55
219	*p*-fluorobenzoic acid	2.07	1.76	1.48
220	phenylacetic acid	1.41	1.83	1.72
221	*m*-chlorophenylacetic acid	2.09	2.13	2.23
222	*p*-chlorophenylacetic acid	2.12	2.12	2.23
223	*m*-bromophenylacetic acid	2.37	2.13	2.42
224	*o*-fluorophenylacetic acid	1.50	2.20	1.93
225	*m*-fluorophenylacetic acid	1.65	2.13	1.91
226	*p*-fluorophenylacetic acid	1.55	2.12	1.89
227	β-phenylpropionic acid	1.84	2.21	2.27
228	4-phenylbutyric acid	2.42	2.51	2.61
229	1-naphthoic acid	3.10	2.79	2.76
230	naphthalene	3.30	3.59	3.36
231	1-methylnaphthalene	3.87	4.07	3.91
232	2-methylnaphthalene	3.86	4.03	3.83
233	1,3-dimethylnaphthalene	4.42	4.53	4.39
234	1,4-dimethylnaphthalene	4.37	4.56	4.40
235	1,5-dimethylnaphthalene	4.38	4.46	4.28
236	2,3-dimethylnaphthalene	4.40	4.50	4.37
237	2,6-dimethylnaphthalene	4.31	4.46	4.32
238	1-nitronaphthalene	3.19	2.95	2.82
239	anthracene	4.45	4.80	4.59
240	9-methylanthracene	5.07	5.15	5.03
241	phenanthracene	4.46	4.83	4.66
242	pyrene	4.88	5.49	5.24
243	fluorene	4.18	3.69	3.77
244	acenaphthene	3.92	3.68	3.69
245	quinoline	2.03	2.69	2.60
246	isoquinoline	2.08	2.69	2.60
247	2,2′-biquinoline	4.31	4.49	4.63
248	biphenyl	4.09	4.10	3.97
249	2-chlorobiphenyl	4.38	4.27	4.24
250	2,4′-dichlorobiphenyl	5.10	4.56	4.62
251	2,5-PCB	5.16	4.59	4.68
252	2,6-PCB	4.93	4.67	4.72
253	2,4,6-PCB	5.47	4.99	5.09
254	bibenzyl	4.79	4.55	4.52

4.5.1.3. Mutagenicity (ln*R*)

The set of compounds used to model mutagenic potency consisted of 95 aromatic and heteroaromatic amines available from the literature.[124] A list of these chemicals and their mutagenic potency is presented in Table 6. The mutagenic potency of the aromatic amines in *S. typhimurium* TA98 + S9 microsomal preparation is expressed by the natural logarithm of the number of revertants per nanomole.

Table 6. Mutagenicity (ln*R*)[a] of 95 Aromatic and Heteroaromatic Amines and Predicted Mutagenicity by Equations (48) and (49)

			Predicted ln*R*	
No.	Chemical name	Obsd. ln*R*	Eq. (48)	Eq. (49)
1	2-bromo-7-aminofluorene	2.62	2.14	2.66
2	2-methoxy-5-methylaniline	−2.05	−2.57	−2.07
3	5-aminoquinoline	−2.00	−1.60	−1.71
4	4-ethoxyaniline	−2.30	−3.75	−3.40
5	1-aminonaphthalene	−0.60	−0.93	−0.86
6	4-aminofluorene	1.13	0.73	1.02
7	2-aminoanthracene	2.62	1.26	1.22
8	7-aminofluoranthene	2.88	1.60	2.27
9	8-aminoquinoline	−1.14	−1.79	−2.02
10	1,7-diaminophenazine	0.75	0.18	0.23
11	2-aminonaphthalene	−0.67	0.21	−0.42
12	4-aminopyrene	3.16	2.99	2.89
13	3-amino-3′-nitrobiphenyl	−0.55	−0.26	0.19
14	2,4,5-trimethylaniline	−1.32	−1.20	−0.55
15	3-aminofluorene	0.89	1.35	1.37
16	3,3′-dichlorobenzidine	0.81	0.24	0.95
17	2,4-dimethylaniline	−2.22	−2.88	−2.34
18	2,7-diaminofluorene	0.48	0.85	1.02
19	3-aminofluoranthene	3.31	2.88	3.06
20	2-aminofluorene	1.93	1.75	1.74
21	2-amino-4′-nitrobiphenyl	−0.62	0.06	0.20
22	4-aminobiphenyl	−0.14	0.10	−0.04
23	3-methoxy-4-methylaniline	−1.96	−3.21	−2.55
24	2-aminocarbazole	0.60	0.61	0.25
25	2-amino-5-nitrophenol	−2.52	−2.65	−3.16
26	2,2′-diaminobiphenyl	−1.52	−0.42	−0.46
27	2-hydroxy-7-aminofluorene	0.41	1.29	1.44
28	1-aminophenanthrene	2.38	1.06	1.19
29	2,5-dimethylaniline	−2.40	−2.41	−2.34
30	4-amino-2′-nitrobiphenyl	−0.92	0.06	0.09
31	2-amino-4-methylphenol	−2.10	−3.59	−2.88
32	2-aminophenazine	0.55	0.83	0.66

(continued)

Table 6. (Continued)

No.	Chemical name	Obsd. ln$R$	Predicted ln$R$ Eq. (48)	Predicted ln$R$ Eq. (49)
33	4-aminophenyl sulfide	0.31	0.32	0.18
34	2,4-dinitroaniline	−2.00	−0.59	−1.54
35	2,4-diaminoisopropylbenzene	−3.00	−1.79	−1.35
36	2,4-difluoroaniline	−2.70	−1.95	−2.68
37	4,4′-methylenedianiline	−1.60	−1.23	−1.19
38	3,3′-dimethylbenzidine	0.01	−0.65	−0.76
39	2-aminofluoranthene	3.23	2.51	2.79
40	2-amino-3′-nitrobiphenyl	−0.89	−0.33	0.04
41	1-aminofluoranthene	3.35	2.61	2.92
42	4,4′-ethylenebis(aniline)	−2.15	−1.79	−1.59
43	4-chloroaniline	−2.52	−2.54	−2.52
44	2-aminophenanthrene	2.46	2.02	1.59
45	4-fluoroaniline	−3.32	−2.85	−3.01
46	9-aminophenanthrene	2.98	1.33	1.26
47	3,3′-diaminobiphenyl	−1.30	−1.28	−1.14
48	2-aminopyrene	3.50	3.43	3.01
49	2,6-dichloro-1,4-phenylenediamine	−0.69	−1.50	−1.78
50	2-amino-7-acetamidofluorene	1.18	1.09	1.46
51	2,8-diaminophenazine	1.12	0.17	0.34
52	6-aminoquinoline	−2.67	−1.31	−1.51
53	4-methoxy-2-methylaniline	−3.00	−3.07	−2.55
54	3-amino-2′-nitrobiphenyl	−1.30	−0.22	−0.10
55	2,4′-diaminobiphenyl	−0.92	0.08	−0.31
56	1,6-diaminophenazine	0.20	0.41	0.54
57	4-aminophenyl disulfide	−1.03	−0.12	0.43
58	2-bromo-4,6-dinitroaniline	−0.54	−1.05	−1.33
59	2,4-diamino-*n*-butylbenzene	−2.70	−2.93	−3.09
60	4-aminophenyl ether	−1.14	−0.50	−0.56
61	2-aminobiphenyl	−1.49	0.02	−0.40
62	1,9-diaminophenazine	0.04	0.13	0.26
63	1-aminofluorene	0.43	0.99	1.11
64	8-aminofluoranthene	3.80	2.77	3.01
65	2-chloroaniline	−3.00	−3.16	−2.95
66	3-amino-ααα-trifluorotoluene	−0.80	−0.78	−1.16
67	2-amino-1-nitronaphthalene	−1.17	−0.24	−0.47
68	3-amino-4′-nitrobiphenyl	0.69	−0.29	0.28
69	4-bromoaniline	−2.70	−2.23	−2.04
70	2-amino-4-chlorophenol	−3.00	−2.56	−2.68
71	3,3′-dimethoxybenzidine	0.15	−0.50	−0.43
72	4-cyclohexylaniline	−1.24	−1.97	−1.89
73	4-phenoxyaniline	0.38	−0.28	−0.80
74	4,4′-methylenebis(*o*-ethylaniline)	−0.99	−0.11	−1.62
75	2-amino-7-nitrofluorene	3.00	1.56	2.36
76	Benzidine	−0.39	−0.98	−0.94

Table 6. (Continued)

No.	Chemical name	Obsd. ln*R*	Predicted ln*R* Eq. (48)	Eq. (49)
77	1-amino-4-nitronaphthalene	−1.77	−0.21	−0.36
78	4-amino-3′-nitrobiphenyl	1.02	−0.36	0.08
79	4-amino-4′-nitrobiphenyl	1.04	−0.27	0.28
80	1-aminophenazine	−0.01	0.67	0.51
81	4,4′-methylenebis(*o*-fluoroaniline)	0.23	0.46	0.66
82	4-chloro-2-nitroaniline	−2.22	−2.31	−2.79
83	3-aminoquinoline	−3.14	−1.50	−1.82
84	3-aminocarbazole	−0.48	0.84	0.51
85	4-chloro-1,2-phenylenediamine	−0.49	−1.50	−1.68
86	3-aminophenanthrene	3.77	1.64	1.30
87	3,4′-diaminobiphenyl	0.20	−0.74	−1.03
88	1-aminoanthracene	1.18	1.32	1.33
89	1-aminocarbazole	−1.04	0.14	−0.11
90	9-aminoanthracene	0.87	1.65	1.45
91	4-aminocarbazole	−1.42	0.33	0.06
92	6-aminochrysene	1.83	2.49	3.31
93	1-aminopyrene	1.43	2.91	2.88
94	4,4′-methylenebis(*o*-isopropylaniline)	−1.77	0.71	−1.29
95	2,7-diaminophenazine	3.97	1.42	1.10

[a] ln*R* = log revertants per nanomole, *S. typhimurium* TA98 with metabolic activation.

4.5.2. Calculation of Parameters

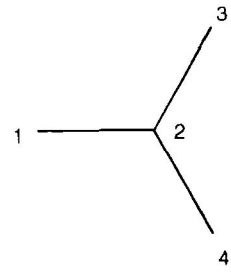
4.5.2.1. Computation of Topological Indexes

The first TI reported in the chemical literature, the Wiener index *W*,[87] may be calculated from the distance matrix $D(G)$ of a hydrogen-suppressed chemical graph G as the sum of the entries in the upper triangular distance submatrix. The distance matrix $D(G)$ of a nondirected graph G with n vertices is a symmetric $n \times n$ matrix (d_{ij}), where d_{ij} is equal to the topological distance between vertices v_i and v_j in G. Each diagonal element d_{ii} of $D(G)$ is zero. We give below the distance matrix $D(G_1)$ of the labeled hydrogen-suppressed graph G_1 of isobutane (Figure 5):

$$D(G_1) = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} (1) & (2) & (3) & (4) \end{array} \\ \begin{bmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{bmatrix} \end{array}$$

W is calculated as:

$$W = \tfrac{1}{2} \sum_{i,j} d_{ij} = \sum_{h} h \cdot g_h \tag{38}$$



G_1

Figure 5. Labeled hydrogen-suppressed graph of isobutane.

where g_h is the number of unordered pairs of vertices whose distance is h.

Randić's connectivity index,[35] higher-order connectivity indexes, and path, cluster, path-cluster, and chain types of simple, bond and valence connectivity parameters developed by Kier and Hall[77] were calculated by a computer program POLLY 2.3 developed by Basak, Harriss, and Magnuson[125] at the University of Minnesota. Also, P_h parameters, the number of paths of length $h(h = 0–10)$ in the hydrogen-suppressed graph, are calculated using standard algorithms.

Balaban[22–24] defined a series of indexes based on distance sums within the distance matrix for a molecular graph which he designated as J indexes. Unlike W, these indexes are independent of molecular size and have low degeneracy.

Information-theoretic TIs are calculated by the application of information theory on molecular graphs. An appropriate set A of n elements is derived from a molecular graph G depending on certain structural characteristics. On the basis of an equivalence relation defined on A, the set A is partitioned into disjoint subsets A_i of order n_i $(i = 1, 2, \ldots, h; \Sigma_i n_i = n)$. A probability distribution is then assigned to the set of equivalence classes:

$$A_1, A_2, \ldots, A_h$$

$$p_1, p_2, \ldots, p_h$$

where $p_i = n_i/n$ is the probability that a randomly selected element of A will occur in the i^{th} subset.

The mean information content of an element of A is defined by Shannon's relation[126]:

$$\text{IC} = -\sum_{i=1}^{h} p_i \log_2 p_i \tag{39}$$

The logarithm base 2 is used to measure the information content in bits. The total information content of the set A is then n times IC.

To account for the chemical nature of vertices, as well as their bonding pattern, Sarkar, Roy, and Sarkar[127] calculated information content of molecular graphs on the basis of an equivalence relation where two atoms of the same element are considered equivalent if they possess an identical first-order topological neighborhood. Since properties of atoms or reaction centers are often modulated by physicochemical characteristics of distant neighbors, i.e., neighbors of neighbors, it was deemed essential to extend this approach to account for higher-order neighbors of vertices. This can be accomplished by defining open spheres for all vertices of a molecular graph. If r is any nonnegative real number, and v is a vertex of the graph G, then the open sphere $S(v,r)$ is defined as the set consisting of all vertices v_i in G such that $d(v,v_i) < r$. Obviously, $S(v,0) = \phi$, $S(v,r) = v$ for $0 < r < 1$, and if $1 < r < 2$, then $S(v,r)$ is the set consisting of v and all vertices v_i of G situated at unit distance from v.

One can construct such open spheres for higher integral values of r. For a particular value of r, the collection of all such open spheres $S(v,r)$, where v runs over the whole vertex set V, forms a neighborhood system of the vertices of G. A suitably defined equivalence relation can then partition Vinto disjoint subsets consisting of topological neighborhoods of vertices up to r^{th} order neighbors. Such an approach has already been developed and the information-theoretic indexes calculated are called indexes of neighborhood symmetry.[128]

In this method, chemical species are symbolized by weighted linear graphs. Two vertices u_0 and v_0 of a molecular graph are said to be equivalent with respect to r^{th} order neighborhood if and only if corresponding to each path $u_0, u_1, \ldots, u_r$ of length r, there is a distinct path $v_0, v_1, \ldots, v_r$ of the same length such that the paths have similar edge weights, and both u_0 and v_0 are connected to the same number and type of atoms up to the r^{th} order bonded neighbors. The detailed equivalence relation is described in our earlier studies.[128]

Once partitioning of the vertex set for a particular order of neighborhood is completed, IC_r is calculated by equation (39). Basak, Roy, and Ghosh[129] defined another information-theoretic measure, structural information content (SIC_r), which is calculated as:

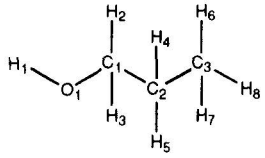
$$SIC_r = IC_r/\log_2 n \tag{40}$$

where IC_r is calculated from equation (39) and n is the total number of vertices of the graph.

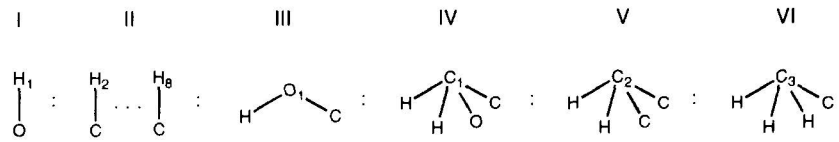
Another information-theoretic invariant, complementary information content (CIC_r), is defined as[81]:

$$CIC_r = \log_2 n - IC_r \tag{41}$$

CIC_r represents the difference between maximum possible complexity of a graph (where each vertex belongs to a separate equivalence class) and the realized topological information of a chemical species as defined by IC_r. Figure 6 provides an example of the first order ($r = 1$) calculations of IC, SIC, and CIC.



First order neighbors:



Subsets:

I	II	III	IV	V	VI
(H_1)	(H_2-H_8)	(O_1)	(C_1)	(C_2)	(C_3)

Probability:

I	II	III	IV	V	VI
1/12	7/12	1/12	1/12	1/12	1/12

$IC_1 = 5*1/12*\log_2 12 + 7/12*\log_2 12/7 = 1.950$ bits
$SIC_1 = IC_1/\log_2 12 = 0.544$ bits
$CIC_1 = \log_2 12 - IC_1 = 1.635$ bits

Figure 6. Derivation of first-order neighborhoods and calculation of complexity indexes (IC_1, SIC_1, and CIC_1) for *n*-propanol.

The information-theoretic index on graph distance, I_D^W is calculated from the distance matrix D (G) of a molecular graph G as follows[34]:

$$I_D^W = W \log_2 W - \sum_h g_h \cdot h \log_2 h$$

The mean information index, $\overline{I}_D^W$, is found by dividing the information index I_D^W by W. IC_r, SIC_r, CIC_r, I_D^W, and $\overline{I}_D^W$ were calculated by POLLY 2.3.[125] The information-

theoretic parameters defined on the distance matrix, H^D and H^V, were calculated by the method of Raychaudhury *et al.*[36]

4.5.2.2. Computation of Geometrical Parameters

Volume (V_W) was calculated using the SYBYL[130] package from Tripos Associates, Inc. The 3D Wiener numbers were calculated using SYBYL with an SPL (Sybyl Programming Language) program. Calculation of 3D Wiener numbers consists of the sum of entries in the upper triangular submatrix of the topographic Euclidean distance matrix for a molecule. The 3D coordinates for the atoms were determined using CONCORD 3.0.1.[131] Two variants of the 3D Wiener number were calculated. For ${}^{3D}W_H$, hydrogen atoms are included in the computations and for ${}^{3D}W$, hydrogen atoms are excluded from the computations.

4.5.2.3. Computation of HB_1

The hydrogen bonding parameter HB_1 was calculated using a program developed by Basak.[132] This program is based on the ideas of Ou *et al.*[133]

The list of parameters used in this chapter is given in Table 2.

4.5.3. Statistical Methods

4.5.3.1. Index Selection

Since the scale of the TIs vary by several orders of magnitude, each TI was transformed by the natural log of the index plus one.

The large number of TIs, and the fact that many of them are highly correlated, confounds the development of predictive models. Therefore, we attempted to reduce the number of TIs to a smaller set of relatively independent variables. Variable clustering[134] was used to divide the TIs into disjoint subsets (clusters) that are essentially unidimensional. These clusters form new variables which are the first principal component derived from the members of the cluster. From each cluster of indexes, a single index was selected. The index chosen was the one most correlated with the cluster variable. In some cases, a member of a cluster showed poor group membership relative to the other members of the cluster, i.e., the correlation of an index with the cluster variable was much lower than the other members. Any variable showing poor cluster membership was selected for further studies as well. A correlation of a TI with the cluster variable less than 0.7 was used as the definition of poor cluster membership.

4.5.3.2. Regression Analysis

The variables used to model each of the properties in this study were TIs, HB_1 and three geometry-related parameters, volume (V_W) and the two 3D Wiener numbers

(^{3D}W and $^{3D}W_H$). The TIs were restricted to those selected by the variable clustering procedure described previously.

All subsets regression was used for the development of the models. The criteria used for defining the "best" model were R^2 and Mallow's C^P.[135] For each of the properties examined, initial models used only the TIs and HB_1 as potential variables. Subsequently, we added the three geometric variables to examine the improvement provided by the addition of geometric information.

4.5.4. Results

4.5.4.1. Boiling Point

HB_1 is zero for all hydrocarbons and, therefore, was deleted from analyses of BP. Twelve of the TIs were deleted for the analysis of the 140 hydrocarbons as well. These indexes included the third- and fourth-order chain connectivity indexes, which were zero for all chemicals, the fourth- and sixth-order bond and valence corrected cluster connectivity indexes, which were perfectly correlated with the simple cluster connectivity indexes ($r = 1.0$), and J^X and J^Y, which were perfectly correlated with J^B for hydrocarbons.

Variable clustering of the remaining 89 TIs resulted in ten clusters. These clusters explained 89.7% of the total variation. In Table 7, we present the indexes selected from each cluster for subsequent use in modeling the BP of hydrocarbons. O, IC_0, IC_1, IC_2, SIC_0, and SIC_1 were selected because of their poor relationship with their clusters ($r < 0.7$).

With the 16 TIs, all subsets regression resulted in a seven-parameter model as follows:

$$BP = -322.86 + 5.46(P_4) - 45.76(IC_1) - 53.23(IC_2) + 799.94(^6\chi_{Ch}) +$$

Table 7. Topological Indexes Selected by Variable Clustering 89 Indexes for the Set of 140 Hydrocarbons with Measured Boiling Point

Cluster	Indices selected	Correlation
1	P_4	0.992
2	CIC_4, IC_0	0.968, 0.338
3	IC_5, O, IC_2, SIC_0	0.969, 0.400, 0.581, 0.456
4	$^5\chi^v_C$	0.964
5	$^5\chi^v_{Ch}$	0.998
6	$^0\chi^v$, SIC_1	0.986, 0.537
7	$^3\chi^v_C$	0.990
8	$^6\chi_{Ch}$, IC_1	0.916, 0.292
9	CIC_2	0.986
10	$^5\chi^v_{PC}$	0.973

(43)
$$288.26(^{0}\chi^{v}) - 32.76(^{3}\chi^{v}_{C}) - 2518.52(^{5}\chi^{v}_{Ch})$$
$$(N = 140,\ r = 0.9956,\ s = 15.5,\ F = 2114)$$

Two chemicals, acenaphthene and cyclopenta(*cd*)pyrene (Nos. 106 and 117 of Table 4, respectively) had rather large residuals (> 60° C). The Cook's distance[136] for these two chemicals indicated they were influential cases. Given these circumstances, an outlier test[136] was performed and both chemicals had a significant result. After the removal of these chemicals, the following model was developed:

(44)
$$BP = -349.11 - 0.71(P_4) - 31.93(IC_1) - 44.70(IC_2) + 884.75(^{6}\chi_{Ch}) +$$
$$291.24(^{0}\chi^{v}) - 33.10(^{3}\chi^{v}_{C}) - 3327.61(^{5}\chi^{v}_{Ch})$$
$$(N = 138,\ r = 0.9976,\ s = 11.4,\ F = 3876)$$

With the inclusion of the geometric parameters, an eight-parameter model was developed, which included two of the geometric parameters:

(45)
$$BP = -626.4 + 1050.8(SIC_0) - 204.0(SIC_1) - 249.8(^{6}\chi_{Ch}) + 364.0(^{0}\chi^{v}) -$$
$$32.3(^{3}\chi^{v}_{C}) + 833.4(^{5}\chi^{v}_{Ch}) + 20.4(^{3D}W)^{1/2} - 8.0(^{3D}W_{H})^{1/2}$$
$$(N = 140,\ r = 0.9994,\ s = 6.1,\ F = 12246)$$

Table 4 presents the predicted normal BP for the hydrocarbons when using equations (44) and (45).

4.5.4.2. Lipophilicity (log*P*)

Twelve of the TIs were dropped from the study of the log*P* data set. The third- and fourth-order chain connectivity indexes were zero for all chemicals and the fifth-order chain connectivity index was nonzero for only one chemical. The sixth-order cluster connectivity indexes were nonzero for only one compound as well. Therefore, 89 indexes were used for the variable clustering.

There were 12 clusters generated by variable clustering for the 89 TIs used for the log*P* data set. The total variation explained by these clusters was 87.8% of the total. Table 8 presents the indexes selected from each of the clusters. The indexes *O*, SIC_0, *J*, IC_0, IC_1, and SIC_2 showed poor membership ($r < 0.7$) within the clusters and were retained as well.

Using all subsets regression with the selected TIs and HB_1 as independent variables resulted in a nine-parameter model:

(46)
$$\log P = -3.64 + 4.81(P_0) - 0.54(\overline{IC}) - 9.30(IC_0) + 13.65(SIC_0) +$$
$$3.88(SIC_4) - 7.68(^{6}\chi_{Ch}) + 0.63(^{6}\chi^{b}_{PC}) - 1.52(J^{B}) - 0.49(HB_1)$$
$$(N = 254,\ r = 0.912,\ s = 0.56,\ F = 134.1)$$

Table 8. Topological Indexes Selected by Variable Clustering 89 Indexes for the Set of 254 Chemicals with Measured log P

Cluster	Indices selected	Correlation
1	P_0	0.981
2	SIC_4	0.944
3	$^5\chi_C^b$	0.929
4	IC_5, SIC_0, O	0.972, 0.469, 0.681
5	$^6\chi_{PC}^b$	0.976
6	$^4\chi_C$	0.980
7	$^6\chi_{CH}$, J	0.855, 0.406
8	SIC_1, IC_0, IC_1, SIC_2	0.910, 0.503, 0.585, 0.689
9	J^B	0.996
10	$^3\chi_C^b$	0.968
11	$\overline{IC}$	0.843
12	$^4\chi$	0.963

Inclusion of the geometric parameters resulted in the following 11-parameter model:

$$\log P = -12.06 - 0.68(\overline{IC}) - 8.13(IC_0) + 2.25(IC_5) + 12.62(SIC_0) - \quad (47)$$
$$5.65(^6\chi_{Ch}) + 0.66(^6\chi_{PC}^b) - 2.22(J) - 0.37(HB_1) +$$
$$4.23 \log(V_W) + 0.60 \log(^{3D}W) - 0.75 \log(^{3D}W_H)$$
$$(N = 254,\ r = 0.932,\ s = 0.50,\ F = 129.1)$$

Predicted values of log$P$ using equations (46) and (47) are presented in Table 5.

4.5.4.3. Mutagenicity

Twelve TIs were dropped from the analyses of the 95 aromatic and heteroaromatic amines. The indexes dropped included the third- and fourth-order chain connectivity indexes, which were zero for all chemicals, and the fourth- and sixth-order cluster connectivity indexes, which were nonzero for only three compounds.

There were eight clusters generated by variable clustering the 89 TIs used for the aromatic amine data set. The total variation explained by these clusters was 88.6% of the total. In Table 9, we present the indexes selected from each of the clusters. Indexes O, $\overline{IC}$, IC_2, SIC_3, CIC_3, and $^3\chi_C^v$ were retained because of their poor cluster membership ($r < 0.7$).

Using all subsets regression with the selected TIs and HB_1 as independent variables resulted in an eight-parameter model:

$$\ln R = 9.308 + 5.141(\overline{IC}) - 3.018(O) - 23.814(IC_3) - 15.050(SIC_1) +$$

Table 9. Topological Indexes Selected by Variable Clustering 89 Indexes for the Set of 95 Aromatic and Heteroaromatic Amines with Measured lnR^a

Cluster	Index selected	Correlation
1	$^4\chi$	0.982
2	SIC_4, SIC_3, CIC_3	0.969, 0.698, 0.665
3	$^6\chi^b_{PC}$, $^3\chi^v_C$	0.951, 0.631
4	SIC_1, O	0.922, 0.478
5	$^6\chi^b_{Ch}$	0.944
6	P_0	0.990
7	$^4\chi^b_{PC}$	0.919
8	IC_3, $\overline{IC}$, IC_2	0.909, 0.698, 0.537

[a]Natural log of number of revertants per nanomole.

$$41.572(SIC_4) + 2.636(^4\chi) + 3.728(^6\chi^b_{PC}) + 3.018(^3\chi^v_C) \quad (48)$$
$$(N = 95,\ r = 0.872,\ s = 0.98,\ F = 34.2)$$

Addition of the geometric parameters resulted in the following model:

$$\ln R = 15.785 + 3.883(\overline{IC}) - 1.374(O) - 14.152(SIC_1) + 2.878(^4\chi) +$$
$$3.409(^6\chi^b_{PC}) + 4.625(^3\chi^v_C) - 7.867(P_0) - 0.0021(^{3D}W_H) + 0.0096(^{3D}W) \quad (49)$$
$$(N = 95,\ r = 0.893,\ s = 0.91,\ F = 37.2)$$

The predicted mutagenicity values for each of the aromatic amine chemicals from equations (48) and (49) are presented in Table 6.

4.6. DISCUSSION

The objectives of this chapter were to review the utility of TIs and 3D parameters in QSARs as well as to report recent results on the relative effectiveness of TIs versus geometrical parameters in the development of QSARs for estimating properties. A large number of QSAR models summarized here show that graph-theoretic invariants correlate reasonably well with physicochemical, biomedicinal, toxicological, and biochemical properties of diverse congeneric sets of molecules. It is also clear that TIs and substructures have found successful applications in the quantification of molecular similarity, selection of analogues, and molecular similarity-based estimation of properties. Of special interest is the fact that the molecular similarity method developed by Basak *et al.*[2] has been successfully used in the discovery of a novel class of human immunodeficiency virus reverse transcriptase (HIV-RT) inhibitors, showing the utility of such nonempirically based methods in practical drug discovery. Examples of

QSARs using 3D descriptors also demonstrate that geometrical parameters alone can predict properties of congeneric molecules quite satisfactorily.

In this context, it was of interest to compare the capabilities of TIs and 3D descriptors in QSAR analysis. To this end, we reported the QSAR studies developed on three different properties, viz., normal boiling of 140 hydrocarbons, lipophilicity (log P, octanol–water) of a diverse set of 254 chemicals, and mutagenicity (ln$R$) of a group of 95 aromatic and heteroaromatic amines. Results of these QSARs show that TIs contain important structural information sufficient to develop useful predictive models for these properties. However, in the case of BP, the addition of geometrical parameters, viz., ${}^{3D}W$ and ${}^{3D}W_H$, to the list of independent variables resulted in improved models in the sense that, while the TI-based QSARs had two outliers, the addition of geometrical variables gave well-behaved models including all of the hydrocarbons in the set. Also, estimate errors were significantly smaller for the regression equation using geometrical descriptors [equation (45) versus (44)]. This indicates that for BP of hydrocarbons, 3D or geometrical parameters encode some pertinent information relevant to BP which are not included in TIs. For logP and mutagenicity, however, the addition of geometrical descriptors resulted in only a slight improvement in the quality of the QSAR models over those derived from TIs only.

For log P, we also used HB_1, an algorithmically derived hydrogen bonding parameter, in addition to TIs, V_W, ${}^{3D}W$, and ${}^{3D}W_H$. This is because the magnitude of log$P$ of a molecule is known to depend significantly on the strength of hydrogen-bonding ability of solutes with solvents.[60,133] Our earlier studies on the correlation of log using algorithmically derived parameters show that HB_1 is an important parameter in predicting log P.[27,31,45] QSARs of log$P$ reported in this chapter provide evidence that the role of HB_1 cannot be carried out by a combination of TIs and 3D parameters.

In many practical situations of drug design and risk assessment, one has to estimate physical/biomedical/toxicological properties of chemicals without access to any empirical data.[55] Similarity-based models[1–3,28,29,47,50,66,111] and estimated values based on nonempirical structural parameters[25,27,31,32,45,70] are two viable alternatives for deriving property values under such data-poor situations. The QSAR models reported here based on TIs and geometrical parameters may find applications in selecting analogues and in estimating properties of chemicals in such cases.

ACKNOWLEDGMENTS

This chapter is contribution No. 159 from the Center for Water and the Environment of the Natural Resources Research Institute. Research reported in this chapter was supported, in part, by grant F49620-94-1-0401 from the United States Air Force, Exxon Corporation, and the Structure–Activity Relationship Consortium (SARCON) of the Natural Resources Research Institute at the University of Minnesota.

REFERENCES

1. M. A. Johnson, S. C. Basak, and G. Maggiora, *Math. Comput. Modeling 11,* 630 (1988).
2. S. C. Basak, V. R. Magnuson, G. J. Niemi, and R. R. Regal, *Discrete Appl. Math. 19,* 17 (1988).
3. M. S. Lajiness, in: *Computational Chemical Graph Theory* (D. H. Rouvray, ed.), pp. 299–316, Nova, New York (1990).
4. R. E. Carhart, D. H. Smith, and R. Venkalaraghavan, *J. Chem. Inf. Comput. Sci. 25,* 64 (1985).
5. S. C. Basak, C. Raychaudhury, A. B. Roy, and J. J. Ghosh, *Indian J. Pharmacol. 13,* 112 (1981).
6. S. K. Ray, S. C. Basak, C. Raychaudhury, A. B. Roy, and J. J. Ghosh, *Arzneim. Forsch. 32,* 322 (1982).
7. S. C. Basak, S. K. Ray, C. Raychaudhury, A. B. Roy, and J. J. Ghosh, *IRCS Med. Sci. 10,* 145 (1982).
8. A. K. Samanta, S. K. Ray, S. C. Basak, and S. K. Bose, *Arzneim. Forsch. 32,* 1515 (1982).
9. S. K. Ray, S. C. Basak, C. Raychaudhury, A. B. Roy, and J. J. Ghosh, *Indian J. Chem. 20B,* 894 (1981).
10. S. C. Basak, D. P. Gieschen, V. R. Magnuson, and D. K. Harriss, *IRCS Med. Sci. 10,* 619 (1982).
11. S. K. Ray, S. C. Basak, C. Raychaudhury, A. B. Roy, and J. J. Ghosh, *Arzneim. Forsch. 33,* 352 (1983).
12. S. C. Basak, D. P. Gieschen, D. K. Harriss, and V. R. Magnuson, *J. Pharm. Sci. 72,* 934 (1983).
13. S. C. Basak, D.K. Harriss, and V. R. Magnuson, *J. Pharm. Sci.* 75, 429 (1984).
14. A. B. Roy, C. Raychaudhury, S. K. Ray, S. C. Basak, and J. J. Ghosh, in: *Proceedings of the Fourth European Symposium on Chemical Structure–Biological Activity: Quantitative Approaches* (J. C. Deardon, ed.), pp. 75-76, Elsevier, Amsterdam (1983).
15. S. K. Ray, S. Gupta, S. C. Basak, C. Raychaudhury, A. B. Roy, and J. J. Ghosh, *Indian J. Chem. 24B,* 1149 (1985).
16. S. C. Basak, L. J. Monsrud, M. E. Rosen, C. M. Frane, and V. R. Magnuson, *Acta Pharm. Jugosl. 36,* 81 (1986).
17. S. C. Basak, B. D. Gute, and L. R. Drewes, *Pharm. Res. 13,* 775 (1996).
18. S. C. Basak, *Med. Sci. Res. 15,* 605 (1987).
19. S. C. Basak, in: *Proceedings of the NATO Advanced Study Institute (ASI) on Pharmacokinetics,* Erice, Sicily, April 4-17, 1994, Plenum, New York.
20. R. Nilakantan, N. Bauman, and R. Venkataraghavan, *J. Chem. Inf. Comput. Sci. 31.* 527 (1991).
21. D. E. Needham, I. C. Wei, and P. G. Seybold, *J. Am. Chem. Soc. 110.* 4186 (1988).
22. A. T. Balaban, *Chem. Phys. Lett. 89,* 399 (1982).
23. A. T. Balaban, *Pure Appl. Chem. 55,* 199 (1983).
24. A. T. Balaban, *MATCH 21,* 115(1986).
25. A. T. Balaban, S. C. Basak, T. Colburn, and G. D. Grunwald, *J. Chem. Inf. Comput. Sci. 34,* 1118 (1994).
26. A. T. Balaban, *J. Chem. Inf. Comput. Sci. 35,* 339 (1995).
27. S. C. Basak, G. J. Niemi, and G. D. Veith, in: *Computational Chemical Graph Theory* (D. H. Rouvray, ed.), p. 235, Nova, New York (1990).
28. S. C. Basak and G. D. Grunwald, *J. Chem. Inf. Comput. Sci. 35,* 366 (1995).
29. S. C. Basak and G. D. Grunwald, *New J. Chem. 19,* 231 (1995).
30. A. T. Balaban, S. Bertelsen, and S. C. Basak, *Math. Chem. 30,* 55 (1994).
31. S. C. Basak, G. J. Niemi, and G. D. Veith, *J. Math. Chem. 4,* 185 (1990).
32. S. C. Basak, G. J. Niemi, and G. D. Veith, *Math. Comput. Modelling 14,* 511 (1990).
33. S. C. Basak, G. J. Niemi, and G. D. Veith, *J. Math. Chem. 7,* 243 (1991).
34. D. Bonchev and N. Trinajstić, *J. Chem. Phys. 67,* 4517 (1977).
35. M. Randić, *J. Am. Chem. Soc. 97,* 6609 (1975).
36. C. Raychaudhury, S. K. Ray, J. J. Ghosh, A. B. Roy, and S. C. Basak, *J. Comput. Chem. 5,* 581 (1984).
37. D. H. Rouvray and R. B. Pandey, *J. Chem. Phys. 85,* 2288 (1986).
38. D. H. Rouvray, *New Sci. May,* 35 (1993).
39. K. Balasubramanian, *SAR QSAR Environ. Res. 2,* 59 (1994).
40. M, Randić, *Int. J. Quantum Chem. Quant. Biol. Symp. 11,* 137 (1984).
41. S. C. Basak, D. P. Gieschen, and V. R. Magnuson, *Environ. Toxicol. Chem. 3,* 191 (1984).

42. S. C. Basak, C. M. Frane, M. E. Rosen, and V. R. Magnuson, *IRCS Med. Sci. 14*, 848 (1986).
43. S. C. Basak, *Med. Sci. Res. 16*, 281 (1988).
44. G. J. Niemi, S. C. Basak, and G. D. Veith, in: *Envirotech Vienna: Proceedings of the First Conference of the International Society of Environmental Protection* (K. Zirm and J. Mayer, eds.), pp. 57–68. W. B. Druck Gmbh and Co., Reiden, Austria (1989).
45. G. J. Niemi, S. C. Basak, G. D. Veith, and G. D. Grunwald, *Environ. Toxicol. Chem. 11*, 893 (1992).
46. S. C. Basak, S. Bertelsen, and G. D. Grunwald, *Toxicol. Lett. 79*, 239 (1995).
47. S. C. Basak and G. D. Grunwald, *SAR QSAR Environ. Res. 2*, 289 (1994).
48. S. C. Basak and G. D. Grunwald, in: *Proceeding of the XVI International Cancer Congress* (R. S. Rao, M. G. Deo, and L. D. Sanghvi, eds.), pp. 413–416, Monduzzi, Bologna, Italy (1995).
49. S. C. Basak and G. D. Grunwald, *SAR QSAR Environ. Res. 3*, 265 (1995).
50. S. C. Basak and G. D. Grunwald, *Chemosphere 31*, 2529 (1995).
51. T. Colburn, D. Axtell, and S. C. Basak, *Mutat. Res.* (in preparation).
52. S. C. Basak, in: *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology* (W. Karcher and J. Devillers, eds.), pp. 83–103, Kluwer Academic, Dordrecht (1990).
53. S. C. Basak, G. D. Grunwald, G. E. Host, G. J. Niemi, and S. Bradbury, *Environ. Toxicol. Chem.* (in preparation).
54. C. Mansch, *Adv. Pharmacol. Chemother. 13*, 45 (1975).
55. C. M. Auer, J. V. Nabholz, and K. P. Baetcke, *Environ. Health Perspect. 87*, 183 (1990).
56. J. C. Arcos, *Environ. Sci. Technol. 21*, 743 (1987).
57. U. Burkert and N. L. Allinger. *Molecular Mechanics, ACS Monograph 177*, American Chemical Society, Washington, DC (1982).
58. W. G. Richards, *Quantum Pharmacology,* Butterworths, London (1977).
59. A. Verloop, W. Hoogenstraaten, and J. Tipker, in: *Drug Design,* Vol. VII (E. J. Ariens, ed.), pp. 165-207, Academic Press, New York (1976).
60. M. J. Kamlet, R. M. Doherty, G. D. Veith, R. W. Taft, and M. H. Abraham, *Environ. Sci. Technol. 20*, 690 (1986).
61. I. Moriguchi and Y. Kanada, *Chem. Pharm. Bull. 25*, 926 (1977).
62. M. Bunge, *Methods, Models and Matter,* Reidel, Dordrecht (1973).
63. F. Harary, *Graph Theory,* Addison–Wesley, Reading, Massachusetts (1969).
64. N. Trinajstić, *Chemical Graph Theory,* CRC Press, Boca Raton, Florida (1983).
65. I. S. Dmitriev, *Molecules Without Chemical Bonds,* Mir Publishers, Moscow (1981).
66. M. Randić, in: *Concepts and Applications of Molecular Similarity* (M. A. Johnson and G. M. Maggiora, eds.), pp. 77–145, John Wiley & Sons, New York (1990).
67. C. Raychaudhury, S. K. Ray, J. J. Ghosh, A. B. Roy, and S. C. Basak, *J. Comput. Chem. 5*, 581 (1984).
68. A. Sabljic and N. Trinajstić, *Acta Pharm. Jugosl. 31*, 189 (1981).
69. O. Mekenyan, S. Dimitrov, and D. Bonchev, *Eur. Polym. J. 19*, 1185 (1983).
70. A. T. Balaban, N. Joshi, L. B. Kier, and L. H. Hall, *J. Chem. Inf. Comput. Sci. 32*, 233 (1992).
71. M. V. Duudea, O. Minailiuc, and A. T. Balaban, *J. Comput. Chem. 12*, 527 (1991).
72. A. T. Balaban, *Theor. Chim. Acta 53*, 355 (1979).
73. P. A. Filip, T. S. Balaban, and A. T. Balaban, *J. Math. Chem. 1*, 61 (1987).
74. A. T. Balaban and V. Feroiu, *Rep. Mol. Theory 1*, 133 (1990).
75. E. J. Kupchik, *Quant. Struct. Act. Relat. 7*, 57 (1988).
76. L. Pogliani, *J. Phys. Chem. 97*, 6731 (1993).
77. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press, New York (1986).
78. S. C. Basak and G. D. Grunwald, *Math. Modelling Sci. Comput. 2*, 735 (1993).
79. D. Bonchev and N. Trinajstić, *Int. J. Quantum Chem. 12*, 293 (1978).
80. S. C. Basak, B. D. Gute, and S. Ghatak, *J. Chem. Inf. Comput. Sci.* (submitted for publication).
81. S.C. Basak and V. R. Magnuson, *Arzneim. Forsch. 33*, 501 (1983).

82. J. V. Soderman, *CRC Handbook of Identified Carcinogens and Noncarcinogens: Carcinogenicity–Mutagenicity Database,* Vol. I, CRC Press, Boca Raton, Florida (1982).
83. S. C. Basak, C. M. Frane, M. E. Rosen, and V. R. Magnuson, *Med. Sci. Res. 15*, 887 (1987).
84. V. R. Magnuson, D. K. Harriss, and S. C. Basak, in: *Studies in Physical and Theoretical Chemistry* (R. B. King, ed.), pp. 178–191, Elsevier, Amsterdam (1983).
85. F. C. Smeeks and P. C. Jurs, *Theor. Chim. Acta 233*, 111 (1990).
86. Y. Gao and H. Hosoya, *Bull. Chem. Soc. Jpn. 61*, 3093 (1988).
87. H. Wiener, *J. Am. Chem. Soc. 69,* 17 (1947).
88. J. R. Platt, *J. Chem. Phys. 15*, 419 (1947).
89. L. H. Hall and L. B. Kier, *Tetrahedron 33*, 1953 (1977).
90. L. Pogliani, *J. Phys. Chem.* 99, 925 (1995).
91. W. J. Boecklen and G. J. Niemi, *SAR QSAR Environ. Res. 2*, 79 (1994).
92. A. T. Balaban and C. Catana, *SAR QSAR Environ. Res. 2*, 1 (1994).
93. S. P. Gupta and P. Singh, *Bull. Chem. Soc. Jpn. 52*, 2745 (1979).
94. M. Randić, *New J. Chem. 15*, 517 (1991).
95. R. H. Rohrbaugh and P. C. Jurs, *Anal. Chem. 60*, 2249 (1988).
96. D. K. Pal, S. K. Purkayaastha, C. Sengupta, and A. U. De, *Indian J. Chem. 31*, 109 (1992).
97. D. H. Rouvray and W. Tatong, *Int. J. Environ. Stud. 33*, 247 (1989).
98. D. H. Rouvray and W. Tatong, *Z. Naturforsch. 41*, 1238 (1986).
99. G. J. Niemi, R. R. Regal, and G. D. Veith, *ACS Symp. Ser. 292*, 148 (1985).
100. I. Lukovits. *J. Chem. Soc. Perkin Trans. 2*, 1667 (1988).
101. G. J. Niemi, G. D. Veith, R. R. Regal, and D. D. Vaishnav, *Environ. Toxicol. Chem.* 6, 515 (1987).
102. V. K. Gombar and K. Enslein, in: *Applied Mu/tivariate Analysis in SAR and Environmental Studies* (J. Devillers and W. Karcher, eds.), pp. 377–414, Kluwer Academic, Dordrecht (1991).
103. B. W. Blake, K. Enslein, V. K. Gombar, and H. H. Borgstedt, *Mutat. Res. 241*, 261 (1990).
104. T. Okuyama, Y. Miyashita, S. Kanaya, H. Katsumi, S. Sasaki, and M. Randic, *J. Comput. Chem. 9*, 636 (1988).
105. C. L. Wilkins and M. Randić, *Theor. Chim. Acta 58*, 45 (1980).
106. M. Randić and N. Trinajsti , *MATCH 13*, 271 (1982).
107. M. Randić, in: *Molecular Basis of Cancer, Part A; Macromolecular Structure, Carcinogens, and Oncogenes* (R. Rein, ed.), pp. 309–318, Alan R. Liss, New York (1985).
108. S. C. Basak, S. Bertelsen, and G. D. Grunwald, *J. Chem. Inf. Comput. Sci. 34*, 270 (1994).
109. A. Leo and D. Weininger, *CLOGP Version 3.2 User Reference Manual*, Medicinal Chemistry Project, Pomona College. Claremont, California (1984).
110. P. Willett. *J. Chem. Inf. Comput. Sci. 23*, 22 (1983).
111. P. Willett and V. Winterman, *Quant. Struct. Act. Relat. 5*, 18 (1986).
112. P. Willett, in: *Concepts and Applications of Molecular Similarity (M.* A. Johnson and G. M. Maggiora, eds.), pp. 43–63, John Wiley & Sons, New York (1990).
113. G. M. Downs and P. Willett, in: *Applied Multivariate Analysis in SAR and Environmental Studies (*J. Devillers and W. Karcher, eds.), pp 247–279, Kluwer Academic, Dordrecht (1991).
114. P. A. Bath, A. R. Andrew, and P. Willett, *J. Chem. Inf. Comput. Sci. 34*, 141 (1994).
115. R. D. Brown. G. Jones, and P. Willett, *J. Chem. Inf. Comput. Sci. 34*, 63 (1994).
116. S. C. Basak, B. D. Gute, and G. D. Grunwald, *Croat, Chem. Acta 69* (1996) (in press).
117. J. E. Amoore, *Nature 214*, 1095 (1967).
118. M. Charton. *Top. Curr. Chem. 114*, 107 (1983).
119. B. Bogdanov, S. Nikolić, and N. Trinajstić, *J. Math. Chem. 3*, 299 (1989).
120. R. P. Bhatnagar, P. Singh, and S. P. Gupta, *Indian J. Chem. 19B*, 780 (1980).
121. R. D. Cramer III, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc. 110*, 5959 (1988).
122. O. Mekenyan, D. Bonchev, and N. Trinajstić, *Int. J. Quantum Chem. 18*, 369 (1980).
123. W. Karcher, *Spectral Atlas of Polycyclic Aromatic Hydrocarbons,* Vol. 2, pp. 16–19, Kluwer Academic, Dordrecht (1988).

124. A. K. Debnath, G. Debnath, A. J. Shusterman, and C. Hansch, *Environ. Mol. Mutagen. 19*, *37* (1992).
125. S.C. Basak, D. K. Harriss, and V. R. Magnuson, POLLY 2.3, copyright by the University of Minnesota (1988).
126. C. E. Shannon, *Bell Syst. Tech. J.* 27, 379 (1948).
127. R. Sarkar, A. B. Roy, and R. K. Sarkar, *Math. Biosci. 39*, 379 (1978).
128. A. B. Roy, S. C. Basak, D. K. Harriss, and V. R. Magnuson. in: *Mathematical Modelling in Science and Technology* (X. J. R. Avula, R. E. Kalman, A. I. Liapis, and E. Y. Rodin, eds.), pp. 745–750, Pergamon Press, Elmsford, New York (1984).
129. S. C. Basak, A. B. Roy, and J. J. Ghosh, in: *Proceedings of the Second International Conference on Mathematical Modelling,* Vol. II (X. J. R. Avula, R. Bellman, Y. L. Luke, and A. K. Rigler, eds.), pp. 851–856, University of Missouri, Rolla(1980).
130. Tripos Associates, Inc., SYBYLVersion 6.1. Tripos Associates. Inc., St. Louis, Missouri (1994).
131. Tripos Associates, Inc., CONCORD Version 3.0.1, Tripos Associates, Inc., St. Louis, Missouri (1993).
132. S. C. Basak, H-BOND: A Program tor Calculating Hydrogen Bonding Parameter. University of Minnesota. Duluth (1988).
133. Y.-C. Ou, Y. Ouyang, and E. J. Lien, *J. Mol. Sci. 4*, 89 (1986).
134. SAS Institute. Inc., *SAS/STAT User's Guide. Release 6.03 Edition.* SAS Institute, Inc., Cary, North Carolina (1988).
135. R. R. Hocking, *Biometrics 32*, 1 (1976).
136. S. Weisberg, *Applied Linear Regression,* John Wiley & Sons, New York (1980).