Use of Topostructural, Topochemical, and Geometric Parameters in the Prediction of Vapor Pressure: A Hierarchical QSAR Approach

Subhash C. Basak,* Brian D. Gute, and Gregory D. Grunwald

Natural Resources Research Institute, University of Minnesota, Duluth, Duluth, Minnesota 55811

Received December 31, 1996[⊗]

Numerous quantitative structure−activity relationships (QSARs) have been developed using topostructural, topochemical, and geometrical molecular descriptors. However, few systematic studies have been carried out on the relative effectiveness of these three classes of parameters in predicting properties. We have carried out a systematic analysis of the relative utility of the three types of structural descriptors in developing QSAR models for predicting vapor pressure at STP for a set of 476 diverse chemicals. The hierarchical technique has proven to be useful in illuminating the relationships of different types of molecular description information to physicochemical property and is a useful tool for limiting the number of independent variables in linear regression modeling to avoid the problems of chance correlations.

1. INTRODUCTION

A large number of quantitative structure−activity relationship (QSAR) studies have been reported in recent literature using theoretical molecular descriptors in predicting physicochemical, pharmacological, and toxicological properties of molecules.[1−15] Such descriptors comprise graph invariants, geometrical or 3-D parameters, and quantum chemical indices. One of the reasons for the current upsurge of interest is the fact that such descriptors can be derived algorithmically, i.e., can be computed for any molecule, real or hypothetical, using standard software. Both in pharmaceutical drug design and in risk assessment of chemicals, one has to evaluate potential biological effects of chemicals. Evaluation schemes based on property−property correlation paradigms are not very useful in practical situations, because, for most of the candidate structures, the experimental data necessary for proper evaluation are not available. This is especially true for the thousands of chemicals rapidly produced by methods of combinatoric chemistry[16] as well as for the large number of chemicals present in the Toxic Substances Control Act (TSCA) Inventory.[17]

A large number of physicochemical and biological endpoints are necessary for estimating the ecotoxicological fate, transport, and effects of environmental pollutants.[17−19] The vapor pressure of chemicals is important in determining the partitioning of chemicals among different phases once they are released in the environment. Many QSARs have been reported for predicting normal vapor pressure of chemicals. Such studies are usually carried out on small sets of congeneric chemicals. Also, many QSARs use experimental data as inputs in the model. Therefore, it becomes necessary to develop QSARs based on nonempirical parameters which can predict the vapor pressure for a heterogeneous collection of chemicals so that such models are generally applicable. With this end in mind, in the current paper we have carried out a QSAR study of 476 diverse chemicals using three types of nonempirical molecular descriptors.

* All correspondence should be addressed to Dr. Subhash C. Basak, Natural Resources Research Institute, University of Minnesota, Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811.

[⊗] Abstract published in *Advance ACS Abstracts,* June 1, 1997.

2. MATERIALS AND METHODS

2.1. Normal Vapor Pressure Database. Measured values for a subset of the Toxic Substances Control Act (TSCA) Inventory[17] were obtained from the ASTER (Assessment Tools for the Evaluation of Risk) database.[20] This subset consisted of a diverse set of chemicals where vapor pressure (p_{vap}) was measured at 25 °C and over a pressure range of approximately 3−10 000 mmHg. Due to the size of the dataset being used in this study, data for these chemicals will not be listed in this paper. An electronic copy of the data may be obtained by contacting the authors.

2.2. Computation of Topological Indices. The majority of the topological indices (TIs) used in this study have been calculated by the computer program POLLY 2.3.[21] These indices include Wiener index,[22] the molecular connectivity indices developed by Randić and Kier and Hall,[1,23] information theoretic indices defined on distance matrices of graphs,[24,25] and a set of parameters derived on the neighborhood complexity of vertices in hydrogen-filled molecular graphs.[2,26−28] Balaban's J indices[29−31] were calculated using software developed by the authors.

van der Waal's volume (V_w)[32−34] was calculated using Sybyl 6.2.[35] The 3-D Wiener numbers[36] were calculated by Sybyl using an SPL (Sybyl Programming Language) program developed by the authors. Calculation of 3-D Wiener numbers consists of the summation of the entries in the upper triangular submatrix of the topographic Euclidean distance matrix for a molecule. The 3-D coordinates for the atoms were determined using CONCORD 3.2.1.[37] Two variants of the 3-D Wiener number were calculated, ${}^{3D}W_H$ and ${}^{3D}W$, where hydrogen atoms are included and excluded from the computations, respectively.

Table 1 provides a complete listing of all of the topological and geometrical parameters which have been used in this study. The listing includes the symbols used to represent the parameters and brief definitions for each of the parameters.

Two additional parameters were used in modeling normal vapor pressure, HB_1, and dipole moment (μ). HB_1 is a simple hydrogen bonding parameter calculated using a program developed by Basak,[38] which is based on the ideas

S0095-2338(96)00176-X CCC: $14.00 © 1997 American Chemical Society

Table 1. Symbols and Definitions of Topological and Geometrical Parameters

I^{W}_{D}	information index for the magnitudes of distances between all possible pairs of vertices of a graph
$\overline{I^{W}_{D}}$	mean information index for the magnitude of distance
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
I_{ORB}	information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices
O	order of neighborhood when IC_r reaches it maximum value for the hydrogen-filled graph
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 $r^{th}(r = 0-5)$ order neighborhood of vertices in a hydrogen-filled graph
SIC_r	structural information content for rth ($r = 0-5$) order neighborhood of vertices in a hydrogen-filled graph
CIC_r	complmentary information content for rth ($r = 0-5$) order neighborhood of vertices in a hydrogen-filled graph
${}^{h}\chi$	path connectivity index of order $h = 0-6$
${}^{h}\chi_{C}$	cluster connectivity index of order $h = 3-6$
${}^{h}\chi_{PC}$	path-cluster connectivity index of order $h = 4-6$
${}^{h}\chi_{Ch}$	chain connectivity index of order $h = 5, 6$
${}^{h}\chi^{b}$	bond path connectivity index of order $h = 0-6$
${}^{h}\chi^{b}_{C}$	bond cluster connectivity index of order $h = 3-6$
${}^{h}\chi^{b}_{Ch}$	bond chain connectivity index of order $h = 5, 6$
${}^{h}\chi^{b}_{PC}$	bond path-cluster connectivity index of order $h = 4-6$
${}^{h}\chi^{v}$	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 $h = 5, 6$
${}^{h}\chi^{v}_{PC}$	valence path-cluster connectivity index of order $h = 4-6$
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 bond types
J^{X}	Balaban's J index based on relative electronegativities
J^{Y}	Balaban's J index based on relative covalent radii
V_W	van der Waal's volume
${}^{3D}W$	3-D Wiener number for the hydrogen-suppressed geometric distance matrix
${}^{3D}W_H$	3-D Wiener number for the hydrogen-filled geometric distance matrix

of Ou *et al.*[39] Dipole moment was calculated using Sybyl 6.2.[35]

2.3. Data Reduction. The set of 92 TIs was partitioned into two distinct subsets: topostructural indices and topochemical indices. The distinction was made as follows: topostructural indices encode information about the adjacency and distances of atoms (vertices) in molecular structures (graphs) irrespective of the chemical nature of the atoms involved in the bonding or factors like hybridization states of atoms and number of core/valence electrons in individual atoms, while topochemical indices quantify information regarding the topology (connectivity of atoms) as well as specific chemical properties of the atoms comprising a molecule. Topochemical indices are derived from weighted molecular graphs where each vertex (atom) is properly weighted with selected chemical/physical properties. These subsets are shown in Table 2.

The partitioning of the indices left 38 topostructural indices and 54 topochemical indices. At this point no further data reduction is called for, since the ratio of the number of observations in the training set (342) to the total number of variables (92 maximum) falls well within the condition limits suggested by Topliss and Edwards[40] for reducing the probability of spurious correlations even at the more conservative $R^2 \geq 0.7$ level.

Table 2. Classification of Parameters used in Modeling Normal Vapor Pressure [$\log_{10}(p_{vap})$]

topological	topochemical	geometric	other parameters
$\overline{I^{W}_{D}}$	I_{ORB}	V_W	HB_1
I^{W}_{D}	IC_0-IC_5	${}^{3D}W$	μ
W	SIC_0-SIC_5	${}^{3D}W_H$	
I^{D}	CIC_0-CIC_5		
H^{V}	${}^{0}\chi^{b}-{}^{6}\chi^{b}$		
H^{D}	${}^{3}\chi^{b}_{C}-{}^{6}\chi^{b}_{C}$		
$\overline{IC}$	${}^{5}\chi^{b}_{Ch}$ and ${}^{6}\chi^{b}_{Ch}$		
O	${}^{4}\chi^{b}_{PC}-{}^{6}\chi^{b}_{PC}$		
M_1	${}^{0}\chi^{v}-{}^{6}\chi^{v}$		
M_2	${}^{3}\chi^{v}_{C}-{}^{6}\chi^{v}_{C}$		
${}^{0}\chi-{}^{6}\chi$	${}^{5}\chi^{b}_{Ch}$ and ${}^{6}\chi^{b}_{Ch}$		
${}^{3}\chi_{C}-{}^{6}\chi_{C}$	${}^{4}\chi^{b}_{PC}-{}^{6}\chi^{b}_{PC}$		
${}^{5}\chi_{Ch}$ and ${}^{6}\chi_{Ch}$	J^{B}		
${}^{4}\chi_{PC}-{}^{6}\chi_{PC}$	J^{X}		
P_0-P_{10}	J^{Y}		
J			

2.4. Statistical Analysis and Hierarchical QSAR. Initially, all TIs were transformed by the natural logarithm of the index plus one. This was done since the scale of some indices may be several orders of magnitude greater than that of other indices. The geometric parameters were transformed by the natural logarithm of the parameter.

Two regression procedures were used in developing the linear models. When the number of independent variables was high, typically greater than 25, a stepwise regression procedure was used to maximize the improvement of the explained variance (R^2). When the number of independent variables was smaller, all possible subsets regression was used. Models were then optimized to reduce problems of variance inflation and collinearity. Regression modeling was conducted using the REG procedure of the statistical package SAS.[41]

The vapor pressure data (p_{vap}) was split into a training set (342 compounds) and a test set (134 compounds), an approximately 75/25 split. Models were developed using the training set of chemicals and then used to predict the p_{vap} values of the test chemicals. Final models were then developed using the combined training and test set of chemicals.

Five sets of indices were used in model development. These sets were constructed as part of a hierarchical approach to QSAR modeling. The hierarchy begins with the simplest indices, the topostructural. After developing our initial model utilizing the topostructural indices, we increase the level of complexity. To the indices included in the best topostructural model, we add all of the topochemical indices and proceed to model p_{vap} using these parameters. Likewise, the indices included in the best model from this procedure are combined with the geometrical indices and modeling is conducted once again. In addition to this hierarchical approach, models were also constructed using the topochemical indices alone and the geometrical indices alone for purposes of comparison.

3. RESULTS

Stepwise regression analyses for $\log_{10}(p_{vap})$ of the training set of chemicals is summarized in Table 3. As shown in

Table 3. Summary of the Regression Results for the Training Set and the Prediction Results for the Test Set for the Hierarchical Analysis of $\log_{10}(p_{vap})$

	training set ($N = 342$)				test set ($N = 134$)	
parameter class	variables included	F	R^2	s	R^2	s
topostructural	${}^1\chi$, ${}^6\chi_C$, P_9	104.6	48.1	0.56	57.9	0.46
topochemical	SIC_0, SIC_2, SIC_3, CIC_0, CIC_1, ${}^3\chi^b_C$, ${}^1\chi^v$, ${}^5\chi^v$, ${}^3\chi^v_C$, J^Y	126.3	79.2	0.36	85.8	0.27
geometrical	${}^{3D}W$, ${}^{3D}W_H$, V_W	168.9	51.8	0.53	62.2	0.44
topostructural + topochemical	${}^1\chi$, P_9, IC_1, SIC_2, CIC_1, ${}^3\chi^b_C$, ${}^1\chi^v$, ${}^3\chi^v$, ${}^6\chi^v$, ${}^3\chi^v_C$, ${}^5\chi^v_{Ch}$	112.5	80.4	0.35	84.7	0.28
all indices	H^V, SIC_1, SIC_2, CIC_0, CIC_3, ${}^6\chi_C$, ${}^1\chi^v$, ${}^3\chi^v$, ${}^6\chi^v_C$, P_6, P_{10}	117.4	79.6	0.35	84.2	0.28
ttg + HB_1 + μ	${}^1\chi$, P_3, P_9, IC_0, ${}^1\chi^b$, ${}^3\chi^b_C$, ${}^1\chi^v$, ${}^3\chi^v$, ${}^3\chi^v_C$, HB_1	160.8	82.9	0.32	83.1	0.29

Table 3, the topostructural model using three parameters resulted in an explained variance (R^2) of 48.1% and a standard error (s) of 0.56. Addition of the topochemical parameters to the three topostructural parameters led to a significant increase in the effectiveness of the model. The resulting model used 12 parameters, two topostructural and ten topochemical. This model had an R^2 of 80.4% and s of 0.35. All subsets regression of the two topostructural and ten topochemical indices retained thus far and the three geometrical indices resulted in the selection of the same 12 parameter model, thus the geometrical indices did not contribute significantly to model development. Several other models were constructed for comparative purposes. Using topochemical indices only, a ten parameter model was developed which had an R^2 of 79.2% and s of 0.36. A geometrical model was developed which utilized all three geometrical indices and resulted in an R^2 of 51.8% and s of 0.53. Finally, two additional stepwise models were developed. One model simply used all indices for a comparison between a simple stepwise analysis of the data and the results of the hierarchical procedure. This resulted in an 11 parameter model with R^2 of 79.6% and s of 0.35. The second model added two new parameters, HB_1 and μ. We thought that it might be possible to improve our modeling by adding in some other nonempirical parameters which could be important to the determination of normal vapor pressure. We selected the parameters HB_1 and μ, since they would be important in intermolecular interactions which could have a dramatic effect on vapor pressure. To look at the addition of these parameters, we conducted a stepwise regression analysis using all topostructural, topochemical, and geometric indices so that we would be able to optimize our model, just as we had done with the previous models. The addition of these parameters led to the selection of a ten parameter model which included three topostructural indices, nine topochemical indices, and HB_1. This was the best model yet, with an R^2 of 82.9% and s of 0.32.

Application of these six models to the test set of chemicals resulted in comparable R^2 and s; actually all models improved slightly on their predictions of the test set, and these values are also listed in Table 3. Based on these results, we decided that it was pointless to develop further models using only geometrical parameters. Also, based on the findings that the geometrical indices did not contribute significantly to any of the training models, they were dropped from the development of final models for the full set of 476 chemicals. However, even though the topostructural indices did not perform well in modeling vapor pressure by themselves, they will be used in model development since they did contribute significantly to most of the models.

Regression analyses of the combined set of 476 chemicals showed similar results for estimating $\log_{10}(p_{vap})$ as analysis of the training set. Using only the topostructural indices, stepwise regression analysis resulted in a five parameter model to estimate vapor pressure:

$$\log_{10}(p_{vap}) = 4.88 + 0.20(O) - 2.56({}^1\chi) + 0.49({}^4\chi_C) + 0.79({}^6\chi_C) + 0.98(P_{10}) \quad (1)$$

$$n = 476, \quad R^2 = 51.5\%, \quad s = 0.53, \quad F = 99.7$$

Stepwise regression using the five topostructural parameters and all topochemical parameters resulted in the selection of the following seven parameter model:

$$\log_{10}(p_{vap}) = 8.44 - 1.77({}^1\chi) + 1.25(P_{10}) - 5.69(IC_1) + 3.91(IC_2) - 1.24(IC_5) + 1.41({}^3\chi^b_C) - 1.70({}^1\chi^v) \quad (2)$$

$$n = 476, \quad R^2 = 79.3\%, \quad s = 0.34, \quad F = 224.0$$

Only two of the topostructural indices used in eq 1 were retained by the stepwise regression procedure used to produce eq 2: ${}^1\chi$ and P_{10}. The improvement in R^2 was significant, increasing from 51.5% for eq 1 to 79.3% for eq 2. Also, the model error decreased significantly, dropping by 0.19 logarithmic units. Since we have dropped the geometrical indices, this becomes our final hierarchical model.

The stepwise regression analysis of only topochemical parameters resulted in a 12 parameter model:

$$\log_{10}(p_{vap}) = 6.65 - 3.44(IC_0) - 1.33(IC_5) + 3.47(SIC_2) + 0.87(CIC_1) - 0.48({}^4\chi^b) + 1.44({}^3\chi^b_C) - 1.00({}^1\chi^v) - 0.41({}^3\chi^v) - 0.70({}^5\chi^v) - 1.08({}^3\chi^v_C) + 1.42({}^6\chi^v_{Ch}) - 1.23(J^Y) \quad (3)$$

$$n = 476, \quad R^2 = 75.8\%, \quad s = 0.38, \quad F = 120.5$$

This model which is inferior to the topostructural + topochemical model (eq 2), because its variance explained is lower and, more importantly, it requires more independent variables (parameters) to achieve this explanation of variance.

Stepwise regression of all indices resulted in the selection of an 11 parameter model. This approach selected three topostructural indices and eight topochemical indices to arrive at the following model:

$$\log_{10}(p_{vap}) = 7.85 - 2.56(H^V) + 1.17({}^6\chi_C) - 5.01(IC_1) + 3.65(IC_2) - 0.99(IC_5) + 0.51(CIC_1) - 1.54({}^1\chi^v) - 0.36({}^3\chi^v) - 0.36({}^4\chi^v) - 1.40({}^6\chi^v_C) \quad (4)$$

$$n = 476, \quad R^2 = 80.4\%, \quad s = 0.33, \quad F = 173.4$$

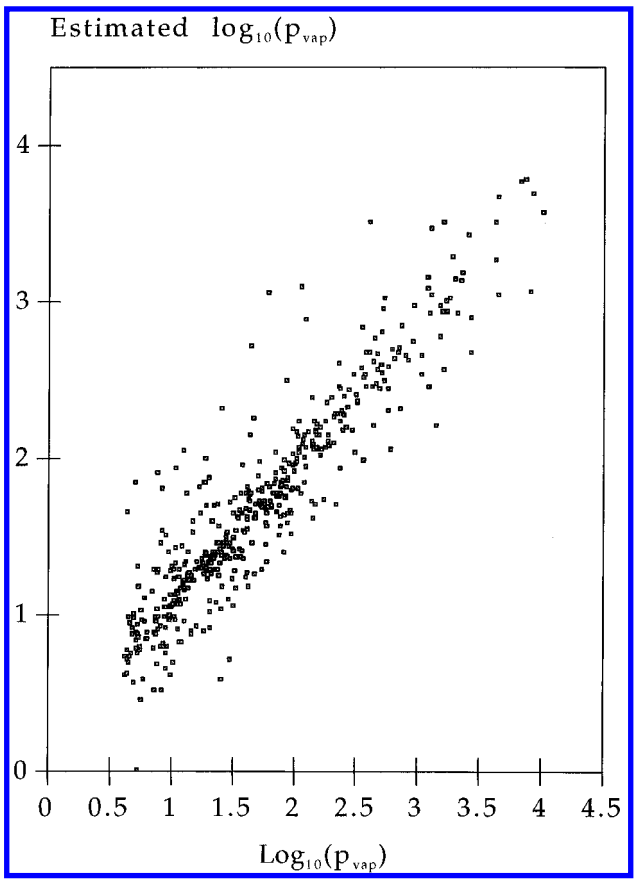

Figure 1. Scatterplot of observed $\log_{10}(p_{vap})$ *vs* estimated $\log_{10}(p_{vap})$ using eq 5 for 476 diverse compounds.

While eq 4 shows some slight improvements over eq 2, the hierarchical model, eq 2 is preferred since it is a simpler model using seven indices instead of 11 and based on a comparison of F values it is a more robust model than that in eq 4.

Finally, we conducted the stepwise regression modeling using all topostructural and topochemical indices with HB_1 and μ for the complete set of 476 chemicals. The resulting ten parameter model used three topostructural indices, six topochemical indices, and HB_1:

$$\log_{10}(p_{vap}) = 9.67 - 3.66(^1\chi) + 0.35(P_3) + 0.74(P_9) - 1.78(IC_0) - 3.33(SIC_1) - 0.81(CIC_2) + 2.05(^2\chi^b) - 1.73(^2\chi^v) - 0.79(^3\chi^v) - 0.29(HB_1) \quad (5)$$

$$n = 476, \quad R^2 = 84.3\%, \quad s = 0.29, \quad F = 249.5$$

Equation 5 shows marked improvement over eq 2, justifying the addition of indices to the model. Also, it meets the criteria on which eq 4 was judged to be lacking. Overall, there is an improvement in variance explained of 5%, with a comparable decrease in standard deviation. A scatter plot of observed $\log_{10}(p_{vap})$ versus estimated $\log_{10}(p_{vap})$ using eq 5 is presented in Figure 1.

4. DISCUSSION

The purpose of this paper was 2-fold: (a) to study the utility of algorithmically-derived molecular descriptors in developing QSAR models for predicting the vapor pressure of chemicals from structure and b) to investigate the relative roles of topostructural, topochemical, and geometrical indices in the estimation of standard vapor pressure.

Table 4. Summary of the Chemical Class Composition of the Normal Vapor Pressure Dataset

compd classification	no. of compds	pure	substituted
total normal vapor pressure dataset	476		
hydrocarbons	253		
non-hydrocarbons[a]	223		
nitro compounds	4	3	1
amines	20	17	3
nitriles	7	6	1
ketones	7	7	0
halogens	100	95	5
anhydrides	1	1	0
esters	18	16	2
carboxylic acids	2	2	0
alcohols	10	6	4
sulfides	39	38	1
thiols	4	4	0
imines	2	2	0
epoxides	1	1	0
aromatic compounds[b]	15	10	4
fused-ring compounds[c]	1	1	0

[a] The non-hydrocarbons are further broken down into the following groups. [b] The 15 aromatic compounds are a mixture of 11 aromatic hydrocarbons and four aromatic halides. [c] The only fused-ring compound was a polycyclic aromatic hydrocarbon.

Results described in this paper (eqs 1−5) show that nonempirical parameters derived predominantly from graph theoretic models of molecules can estimate normal vapor pressure of diverse chemicals reasonably well. The explained variance of data ($R^2 = 84.3\%$) is excellent in view of the fact that the database of chemicals analyzed in this paper is very diverse (see Table 4). It should be mentioned that most published QSAR models for the estimation of vapor pressure have dealt with much smaller data sets with limited structural variety.[42,43]

The relative effectiveness of topostructural, topochemical, and geometrical indices in predicting normal vapor pressure of chemicals is evident from the result presented above. Equation 1 explains over 51% of variance in the data. All parameters used to derive eq 1 are topostructural, *i.e.*, they are parameters which encode information about the adjacency and distance of vertices in skeletal molecular graphs without quantifying any explicit information about such chemical aspects like bond order, electronic character of atoms, etc. Yet, the high explained variance of the property indicates that adjacency and distance in chemical graphs, being general descriptors of molecular size, shape, and branching, are important in predicting properties. This may explain the success of parameters like simple connectivity indices in estimating many diverse properties.[1]

Equation 3 is derived only from topochemical indices. The explained variance of vapor pressure (75.8%) shows that topochemical parameters, as a class, explain a larger fraction of the variance as compared to models derived from only topostructural indices (eq 1). Geometrical parameters were dropped from the set of descriptors after their limited success in prediction for the training and test sets. This is in line with our earlier studies with normal boiling point and hydrophobicity, where it was reported that the addition of geometrical indices could not significantly improve the predictive power of QSAR models derived from a combined set of topostructural and topochemical parameters.[15] It would

be interesting to see whether this pattern holds good for other properties as well. Finally, the addition of the simple nonempirical parameter, HB_1, which contains information relevant to intermolecular interactions further improves our ability to estimate normal vapor pressure resulting in an explained variance of 84.3% (eq 5).

ACKNOWLEDGMENT

This is contribution number 209 from the Center for Water and the Environment of the Natural Resources Research Institute. Research reported in this paper was supported in part by grants F49620-94-1-0401 and F49620-96-1-0330 from the United States Air Force, a grant from Exxon Corporation and the Structure−Activity Relationship Consortium (SARCON) of the Natural Resources Research Institute of the University of Minnesota. The authors would like to dedicate this paper to Professor Milan Randić in appreciation of his contributions in chemical information, quantitative structure−activity relationships, and chemical graph theory.

REFERENCES AND NOTES

(1) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Chichester, England, 1986.
(2) Basak, S. C. Use of Molecular Complexity Indices in Predictive Pharmacology and Toxicology: A QSAR Approach. *Med. Sci. Res.* **1987**, *15*, 605−609.
(3) Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. Correlation Between Structure and Normal Boiling Points of Haloalkanes C_1-C_4 Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118−1121.
(4) Basak, S. C. A Nonempirical Approach to Predicting Molecular Properties Using Graph-Theoretic Invariants. In *Practical Applications of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht/Boston/London, 1990; pp 83−103.
(5) Basak, S. C.; Bertelsen, S.; Grunwald, G. Application of Graph Theoretical Parameters in Quantifying Molecular Similarity and Structure-Activity Studies. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 270−276.
(6) Basak, S. C.; Bertelsen, S.; Grunwald, G. D. Use of Graph Theoretic Parameters in Risk Assessment of Chemicals. *Toxicol. Lett.* **1995**, *79*, 239−250.
(7) Basak, S. C.; Grunwald, G. D. Use of Topological Space and Property Space in Selecting Structural Analogs. *Mathematical Modelling and Scientific Computing*. In press.
(8) Basak, S. C.; Grunwald, G. D. Molecular Similarity and Risk Assessment: Analog Selection and Property Estimation Using Graph Invariants. *SAR and QSAR in Environ. Res.* **1994**, *2*, 289−307.
(9) Basak, S. C.; Grunwald, G. D. Estimation of Lipophilicity From Molecular Structural Similarity. *New J. Chem.* **1995**, *19*, 231−237.
(10) Basak, S. C.; Grunwald, G. D.; Niemi, G. J. Use of Graph-Theoretic and Geometrical Molecular Descriptors in Structure-Activity Relationships. In *From Chemical Topology To Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum Press: New York, 1997; pp 73−116.
(11) Basak, S. C.; Gute, B. D. Use of Graph Theoretic Parameters in Predicting Inhibition of Microsomal Hydroxylation of Anilines by Alcohols: A Molecular Similarity Approach. In *Proceedings of the International Congress on Hazardous Waste: Impact on Human and Ecological Health*; Johnson, B. L., Xintaras, C., Andrews, J. S., Jr. Eds.; Princeton Scientific Publishing Co., Inc.: Princeton, NJ, 1997; pp 492−504.
(12) Basak, S. C.; Gute, B. D.; Drewes, L. R. Predicting Blood-Brain Transport of Drugs: A Computational Approach. *Pharm Res.* **1996**, *13*, 775−778.
(13) Basak, S. C.; Gute, B. D.; Grunwald, G. D. Development and Applications of Molecular Similarity Methods Using Nonempirical Parameters. *Math Modelling Sci. Computing* In press.
(14) Basak, S. C.; Gute, B. D.; Grunwald, G. D. Estimation of Normal Boiling Points of Haloalkanes Using Molecular Similarity. *Croat Chem Acta*. **1996**, *69*, 1159−1173.
(15) Basak, S. C.; Gute, B. D.; Grunwald, G. D. A Comparative Study of Topological and Geometrical Parameters in Estimating Normal Boiling Point and Octanol/Water Partition Coefficient. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1054−1060.
(16) Martin, Y. C. Opportunities for Computational Chemists Afforded by the New Strategies in Drug Discovery: An Opinion. *Network Science* **1996**.
(17) Auer, C. M.; Nabholz, J. V.; Baetcke, K. P. Mode of Action and the Assessment of Chemical Hazards in the Presence of Limited Data: Use of Structure−Activity Relationships (SAR) Under Tsca, Section 5. *Environ Health Perspect.* **1990**, *87*, 183−197.
(18) NRC. *Toxicity Testing: Strategies to Determine Needs and Priorities*; National Academy Press: Washington, DC, 1984; p 84.
(19) Basak, S. C.; Niemi, G. J.; Veith, G. D. Predicting Properties of Molecules Using Graph Invariants. *J. Math. Chem.* **1991**, *7*, 243−272.
(20) Russom, C. L.; Anderson, E. B.; Greenwood, B. E.; Pilli, A. ASTER: An Integration of the AQUIRE Data Base and the QSAR System for Use in Ecological Risk Assessments. *Sci. Total Environ.* **1991**, *109/110*, 667−670.
(21) Basak, S. C.; Harriss, D. K.; Magnuson, V. R. *POLLY Version 2.3*; Copyright of the University of Minnesota, 1988.
(22) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.
(23) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(24) Raychaudhury, C.; Ray, S. K.; Ghosh, J. J.; Roy, A. B.; Basak, S. C. Discrimination of Isomeric Structures Using Information Theoretic Topological Indices. *J. Comput. Chem.* **1984**, *5*, 581−588.
(25) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517−4533.
(26) Basak, S. C.; Magnuson, V. R. Molecular Topology and Narcosis: A Quantitative Structure-Activity Relationship (QSAR) Study of Alcohols Using Complementary Information Content (CIC). *Arzneim. Forsch.* **1983**, *33*, 501−503.
(27) Basak, S. C.; Roy, A. B.; Ghosh, J. J. In *Proceedings of the Second International Conference on Mathematical Modelling*; Avula, X. J. R., Bellman, R., Luke, Y. L., Rigler, A. K., Eds.; University of Missouri−Rolla: Rolla, MO, 1980; p 745.
(28) Roy, A. B.; Basak, S. C.; Harriss, D. K.; Magnuson, V. R. Neighborhood Complexities and Symmetry of Chemical Graphs and Their Biological Applications. In *Mathematical Modelling in Science and Technology*; Avula, X. J. R., Kalman, R. E., Liapis, A. I., Rodin, E. Y., Eds.; Pergamon: New York, 1984; p 745.
(29) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.
(30) Balaban, A. T. Topological Indices Based on Topological Distances in Molecular Graphs. *Pure Appl. Chem.* **1983**, *55*, 199−206.
(31) Balaban, A. T. Chemical Graphs. Part 48. Topological Index *J* for Heteroatom-Containing Molecules Taking into Account Periodicities of Element Properties. *Math. Chem. (MATCH)* **1986**, *21*, 115−122.
(32) Bondi, A. van der Waal's Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.
(33) Moriguchi, I.; Kanada, Y. Use of van der Waal's Volume in Structure−Activity Studies. *Chem. Pharm. Bull.* **1977**, *25*, 926−935.
(34) Moriguchi, I.; Kanada, Y.; Komatsu, K. van der Waal's Volume and the Related Parameters for Hydrophobicity in Structure−Activity Studies. *Chem. Pharm. Bull.* **1976**, *24*, 1799−1806.
(35) Tripos Associates, Inc. *SYBYL Version 6.2*; Tripos Associates, Inc.: St. Louis, MO, 1994.
(36) Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the Three-Dimensional Wiener Number. *J. Math. Chem.* **1989**, *3*, 299−309.
(37) Tripos Associates, Inc. *CONCORD Version 3.2.7*; Tripos Associates, Inc.: St. Louis, MO, 1995.
(38) Basak, S. C. *H-Bond*; Copyright of the University of Minnesota, 1988.
(39) Ou, Y. C.; Ouyang, Y.; Lien, E. J. *J. Mol. Sci.* **1986**, *4*, 89.
(40) Topliss, J. G.; Edwards, R. P. Chance Factor in Studies of Quantitative Structure−Activity Relationships. *J. Med. Chem.* **1979**, *22*, 1238−1244.
(41) SAS Institute Inc. In *SAS/STAT User's Guide, Release 6.03 Edition*. SAS Institute Inc.: Cary, NC, 1988; Chapters 28 and 34, pp 773−875, 949−965.
(42) Drefahl, A.; Reinhard, M. *Handbook for Estimating Physico-Chemical Properties of Organic Compounds*; Stanford University Bookstore, Stanford, CA, 1995.
(43) Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H. *Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Compounds*; McGraw-Hill Book Company: New York, 1982.

CI960176D