

PERSPECTIVE

Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties

Alan R. Katritzky,^{*,†} Uko Maran,[‡] Victor S. Lobanov,[§] and Mati Karelson[‡]

Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P.O. Box 117200, Gainesville, Florida 32611-7200, Department of Chemistry, University of Tartu, Jakobi Street 2, Tartu 51014, Estonia, and 3-Dimensional Pharmaceuticals, Inc., 665 Stockton Drive, Suite 104, Exton, Pennsylvania 19341

Received August 16, 1999

I. INTRODUCTION

Quantitative structure–activity relationship (QSAR) techniques have become indispensable in all aspects of research into the molecular interpretation of biological properties. Today it would be inconceivable for any commercial, governmental, or academic group to research in these fields without the help of sophisticated calculations, as documented in numerous books and reviews.^{1–6}

By contrast, the application of quantitative structure–property relationships (QSPRs) to matters of technological importance has been markedly slower, although QSPR preceded QSAR, and QSPR is usually far less complex and more amenable to physical interpretation than QSAR. Since its origins around the middle of the century, emphasis has been placed on giving QSPR a firm theoretical basis, and classical contributions were made particularly by the groups of Jurs,⁷ Kier and Hall,⁸ Politzer,⁹ Hilal,¹⁰ Abraham,^{11,12} Katritzky and Karelson,¹³ Randic,¹⁴ Balaban,¹⁵ Trinjastic,¹⁶ and others.

Until about 10 years ago, most QSPR correlations dealt with data sets of limited structural diversity, frequently restricted to hydrocarbons. More recently the advent of general programs, particularly ADAPT^{7,17} and CODESSA,^{13,18} has led to the investigation of diverse data sets. This development is likely to continue and to expand to encompass many technologically important areas. With the increasing expense of synthesis, and the increasing emphasis on the creation of more efficient and less costly solutions, the QSPR method is increasingly attractive. In principal, any and every property of a molecule is encoded in its chemical structure, and QSPR frequently offers the best hope to decode it, to understand how structure influences properties, and to enable the prediction and preparation of optimum structures.

Previous reviews of QSPR from our laboratories have included general introductory overviews^{13,19} and a review specifically of the applications of CODESSA.¹⁸ In the present paper we seek to give a general review of the QSPR of diverse data sets. To keep it within reasonable length, work dealing with group contribution methods, and with data sets of limited structural diversity with less than 50 data points,

has been included only when needed for subsequent discussions (references to work of this type can be found in most of our previous publications). Our emphasis is on structurally diversified data sets of potential technological importance. In this paper we do not address descriptor generation,^{8,19–25} statistical methods,^{26–32} or available programs,^{3,18,33} information about which is available from the references quoted.

Available results are classified into those dealing with properties of single molecules, interactions between different molecular species, surfactant properties, and more complex properties, particularly those of polymers. A brief introduction is given to each of these main sections.

II. PROPERTIES OF SINGLE-MOLECULE SPECIES

One of the first attempts to link a property to critical structural features was reported in 1925 when Langmuir proposed linking intermolecular interactions in the liquid state to the surface energy.³⁴ Accordingly, the surface energy of a hydrocarbon was expressed in terms of its total molecular surface area, proceeding from the assumption that the surface energy per unit surface is uniform.³⁵ Subsequently, a whole range of single molecular species properties have been described in terms of structural descriptors.

Boiling Point. The boiling point of a compound is predetermined by the intermolecular interactions in the liquid and by the difference in the molecular internal partition function in the gas phase and in the liquid at the boiling temperature. Therefore, it is expected to be related directly to the chemical structure of the molecule, and indeed numerous methods have been developed over the years for estimating the normal boiling point of a compound from its structure.^{36,37} Pioneering work in applying QSPR to boiling points was done by Wiener who introduced the path number *w* (later named the Wiener index) defined as the sum of the distances between any two carbon atoms in the molecule.³⁸ Since this, there have been very extensive efforts to apply structural information to fit experimental boiling points. Most of this work has concerned homologous and congeneric series of compounds. For detailed overviews see inter alia refs 39 and 40 and, additionally, recent studies on condensed benzenoids,⁴¹ chlorosilanes,⁴³ and alkanes.⁴³ However, non-congeneric/diverse series of compounds are the most challenging.

[†] University of Florida.[‡] University of Tartu.[§] 3-Dimensional Pharmaceuticals, Inc.

Numerous molecular descriptors have been developed to account for inter- and intramolecular interactions in condensed media, on the basis of physical models of different complexity. For instance, Grigoras⁴⁴ estimated the boiling points of diverse organic compounds using the assumption that the dominant intermolecular interaction is related to the molecular surface energy, derived from the molecular surface area and the charge density distribution. The multilinear model thus derived gave a good correlation ($R^2 = 0.958$) with the boiling points of 137 diverse organic compounds. It included four parameters: total molecular surface area, the sums of the positively and negatively charged atomic surface areas multiplied by the corresponding partial charges, and a hydrogen-bonding term. This correlation required atomic charge scaling factors to correct the partial charges calculated by the extended Hückel theory.

For a better representation of the intermolecular interactions, Jurs and Stanton introduced (simultaneously with Grigoras) charged partial surface area (CPSA) descriptors, which combine solvent accessible surface areas with partial atomic charges.⁴⁵ Jurs and co-workers used CPSA descriptors in combination with various constitutional, topological, and other descriptors. The CPSA descriptors were demonstrated to be useful for a homologous series of heterocycles, especially when hydrogen-bonding-specific descriptors were included.^{46,47} For a diverse set of 752 furans, tetrahydrofurans, thiophenes, pyrans, and pyrroles, Jurs and co-workers commenced with a reduced set of 31 descriptors,⁴⁷ which provided an 11-descriptor model ($R^2 = 0.954$, $s = 13.1$ °C) including various connectivity, electronic, constitutional, and CPSA descriptors. These authors concluded that the structural features of the nitrogen heterocycles differed sufficiently from those of the sulfur and oxygen heterocycles and that the same descriptors cannot adequately encode enough information for a combined set. In general, the authors found that at the time it was not possible to produce good-quality predictive equations using large and diverse data sets. The utility of the CPSA descriptors is also evident in a model developed for the boiling points of 298 diverse organic compounds ($R^2 = 0.976$, $s = 11.85$ K), with 8 parameters employed, from which 3 are CPSA descriptors.⁴⁸ Those descriptors provide unique physical insight in that the CPSA descriptors include information on negative charge, fragment surface area of the negatively charged atom, and total molecular surface area and suggest that the accessibility of the heteroatoms play a prominent role in determining the tendency to enter the vapor phase. The topological descriptors employed in the eight-parameter equation help to distinguish among structural isomers. The model was validated with an external data set which shows good agreement ($R^2 = 0.987$, $s = 13.23$) with the training set. In a subsequent study, Jurs and co-workers presented two QSPR models that accurately predict normal boiling points for organic compounds containing heteroatoms.⁴⁹ The first model is for a set of compounds containing only O, S, and halogens as heteroatoms and involves 10 descriptors ($R^2 = 0.982$, $s = 11.6$ K, $n = 248$). The second, for a set of N-containing compounds, also comprises 10 descriptors ($R^2 = 0.960$, $s = 10.7$ K, $n = 90$).

Chemical species are conveniently represented by molecular graphs, with atoms as vertices and bonds as edges. Such a representation provides topological descriptors to encode

the respective structured information.²¹ The combination of topological descriptors with information about compound atomic contributions and properties enables the deviation of topochemical descriptors,⁸ which can be successfully applied to the prediction of boiling points. A diverse set of 1023 compounds, selected from the Toxic Substances Control Act Inventory, was analyzed against topological, topochemical, and geometrical descriptors.⁵⁰ The most successful model ($R^2 = 0.96$, $s = 18.0$ °C) used nine descriptors: two topological and seven topochemical.

Successful preliminary QSPR studies of the boiling points of small groups of compounds^{51,52} encouraged us to model data sets comprising a large structural and functional variability. Our first QSPR treatment of normal boiling points, for a set of 298 structurally variable organic compounds,³⁹ provided a highly significant two-parameter correlation ($R^2 = 0.9544$, $s = 16.2$ K). The two theoretical descriptors involved each had well-defined physical meaning. The first descriptor (gravitation index) is connected with the bulk cohesiveness, dispersion, and cavity-formation effects in liquids. The second descriptor, the hydrogen donor charged surface area, is connected with the hydrogen-bonding ability of the molecule. A more refined QSPR model (with $R^2 = 0.9732$ and $s = 12.4$ K) included, in addition, the most negative atomic partial charge and the number of chlorine atoms in the molecule. The four-parameter equation offered an average prediction error of 2.3% for a standard set of compounds with an average experimental error of 2.1%. In a followup work, the data set of 298 compounds was extended to provide a more diverse and general data set of 584 organic compounds containing C, H, N, O, S, F, Cl, Br, and I atoms.⁴⁰ This set was divided into subsets according to the molecular functionalities, and additional descriptors were sought for each subset, which together with the gravitation index and the charged surface area of hydrogen donor atoms would model the boiling points. A final global eight-parameter correlation model had $R^2 = 0.965$ and a standard error of 15.5 K, which is close to the estimated average experimental error (cf. above). The model appears to be general for a wide variety of organic compounds and expands and refines the conclusions of previous correlation models of the boiling point.

Recently Balaban et al. tested the applicability of various classes of descriptors and multilinear regression (MLR) techniques to develop QSPR models for halogen-, oxygen-, and sulfur-containing compounds.⁵³ The boiling points of 185 compounds containing oxygen or sulfur can be accurately computed with an MLR equation containing six theoretical descriptors, and having the following statistical parameters: $R^2 = 0.984$ and $s = 6.3$ °C. For a set of 534 halogenated alkanes, the best MLR equation with five descriptors has $R^2 = 0.98$ and $s = 9.0$ °C. For subsets of molecules corresponding to monoethers, monosulfides, F- and Cl-containing compounds, or F-containing compounds, even better QSPR models were obtained.

Recent years have witnessed the growth in QSPR applications related to neural networks (NNs), genetic algorithms (GAs), and evolutionary programming (EP). These methods, can help to improve the linear QSPR models by determining nonlinear relationships in those equations and are capable of deriving completely new QSPR models.

Cherqaoui and Villemin used back-propagation NNs (10:7:1 architecture) to find relationships between the structure and boiling point of 150 alkanes. The alkanes were represented by a 10-number structural code, with each digit indicating the number of C–C bonds in which each atom participates, starting from the first chain atom. The three-layer network resulted in good correlation between predicted and experimental boiling points ($R^2 = 0.9954$) and was slightly better in performance compared with the MLR methods.⁵⁴ Cherqaoui et al. also applied a combination of embedding frequency descriptors and NNs to predict boiling points of 185 acyclic ethers, peroxides, acetals, and their sulfur analogues.⁵⁵ The 20:5:1 architecture resulted in a good correlation coefficient ($R^2 = 0.996$) and was slightly better in comparison with the 15-descriptor MLR equation ($R^2 = 0.984$) from the same publication and also slightly better than the 3-descriptor equation ($R^2 = 0.972$) published elsewhere.⁵⁶

Subsequently to their conventional QSPR treatment, Jurs and co-workers applied neural networks to model various classes of compounds: (i) a combined set of furans, tetrahydrofurans, thiophenes, pyrans,⁵⁷ (ii) pyridines,⁵⁷ and (iii) a diverse set of organic compounds.⁴⁸ For the first set they used a network architecture of 16:3:1 which had 55 adjustable parameters for the training set of 270 compounds.⁵⁷ The training of networks resulted in root mean square (RMS) error $s = 8.49$ °C for the cross-validation set, compared to $s = 11.8$ °C for the conventional QSPR methods. The set of 300 pyridines was trained using 7:3:1 NN architecture.⁵⁷ The conventional QSPR and NN treatments had the same quality of prediction for the training set, $s = 14.97$ °C and $s = 14.57$ °C, respectively. For the cross-validation set, the NNs outperformed conventional QSPR in the case of the pyridines with $s = 16.50$ °C and $s = 19.63$ °C, respectively.⁵⁷ For the analysis of a diverse set of organic compounds, Jurs and co-workers used back-propagation with the BFGS (Broyden–Fletcher–Goldford–Shannon) quasi-Newton optimization method with 8:3:1 architecture for NNs on a training set of 241 compounds.⁴⁸ The back-propagation NN combination with the BFGS method resulted in 1 K improvement over the MLR techniques, with the RMS errors 11.18 K for the training set and 9.17 K for the cross-validation set.

Hu and co-workers applied two different sets of five topological descriptors to model the boiling points of alkenes with a 5:5:1 architecture of NNs.^{58,59} Sets of 51 and 49 compounds were used to produce a trained network, which predicted the boiling points in the test set with relative standard derivations (RSDs) of 2.3%⁵⁸ and 3.08%,⁵⁹ respectively.

Hall and Story analyzed the normal boiling points of a set of 298 heterogeneous organic compounds with NN 19:5:1 architecture with electrotopological state indices.⁶⁰ This model produced a mean absolute error of 3.93 K. The conventional QSPR models have a lower predicting quality, but include a lower number of descriptors, eight ($s = 11.85$ K)⁴⁸ and four ($s = 12.4$ K),³⁹ respectively.

Wessel and Jurs also used NNs to improve the prediction of normal boiling points of hydrocarbons based on six important descriptors obtained from the respective MLR equation.⁶¹ The 6:5:1 architecture gave an RMS of 5.7 K for the training set of 267 compounds, which was slightly

better than the respective root mean square for the MLR equation ($s = 6.3$ K).

Gakh et al. used the back-propagation-type network together with graph theory to derive the descriptors for modeling of structure–property relationships on boiling points of hydrocarbons. The network was trained with 106 compounds and 7 descriptors, yielding the quality of prediction with an average error of 1.19% on the prediction set of 25 hydrocarbons.⁶²

Critical Temperature. Critical temperature is an important technological property, which is determined by a multiplicity of intermolecular interactions in the liquid state. As established by the principle of corresponding states, all substances obey the same equation of state in terms of the reduced variables temperature, pressure, and volume scaled, respectively, by the critical temperature, critical pressure, and critical volume. The reduced variables characterize a specific state of a substance. Therefore, the critical properties, which are the scaling factors, express the individual features of each substance, and their variation is determined by the molecular structural differences between substances. Several groups have studied the predictability of critical temperature from molecular structure based on small sets of compounds.⁶³ Here we concentrate on large and diverse sets.

Grigoras has applied electrostatic molecular surface interaction descriptors in a QSPR study of the critical temperatures of 137 organic molecules.⁴⁴ These descriptors were especially designed to account for the polar interactions of various heteroatoms. To account for these interactions, the molecular surface is divided into atomic surface contributions, accounting for dispersion, polar, and hydrogen-bonding interactions. A four-descriptor MLR equation accounted for (i) dispersion forces which were defined by total molecular surface area, (ii) polar (electrostatic) interactions as expressed by an electrostatic negative or positive surface interaction term, and (iii) hydrogen bond interactions with a hydrogen-bonding molecular surface interaction term. The model was of good quality ($R^2 = 0.964$); the standard deviation of 16.9 °C corresponded to about 2.9% of the mean value for the critical temperatures listed.

Jurs and co-workers used an MLR equation with eight descriptors to correlate the critical temperatures of 147 diverse organic molecules with $R^2 = 0.978$ and $s = 11.9$ K.⁴⁸ Two CPSA descriptors together with simple counts of atoms and bonds, topological descriptors, and charge distribution were included. Recently they repeated the QSPR analysis of the same data set, with additional descriptors and using simulated annealing and genetic algorithms as new methods for descriptor selection.⁶⁴ The test set of 147 compounds (2 compounds were omitted as statistical outliers) gave $R^2 = 0.986$ and $s = 9.16$ K with 8 descriptors, which is in good agreement with the experimental error (8 K) reported for the original database. The test set of 17 compounds (acetone was omitted) showed a similar quality of prediction ($s = 9.13$ K).

Politzer and his group used theoretical descriptors based on electrostatic potential⁶⁵ to correlate critical temperatures.⁶⁶ Electrostatic potential is particularly well suited for studies of intermolecular interactions, e.g., hydrogen-bonding tendencies and solute/solvent (or solute/solute) and recognition interactions, which can be reviewed as being determined largely by electrostatic factors. The “total surface potential

variance" (σ_{tot}^2) and "balance" (ν) parameters were introduced. The total variance σ_{tot}^2 measures the range of the surface potential and is particularly sensitive to variation in its magnitude, emphasizing positive and negative extremes. It can be interpreted as an indicator or the tendency of molecules to interact electrostatically. The balance parameter ν helps to represent accurately the manner in which the total variance affects interactive tendencies. This study used ab initio self-consistent molecular orbital calculations. The two-descriptor MLR model obtained for the critical temperature ($R^2 = 0.835$, $n = 66$, $s = 59.7$ K) included as a descriptor the square root of the cross-term for $\sigma_{\text{tot}}^2\nu$.

The large descriptor sets and methodology implemented in CODESSA software enabled the successful development of a three-parameter MLR model for critical temperatures within a set of 165 structurally diverse molecules⁶³ with an average error of 16.8 K (with $R^2 = 0.955$, $R_{\text{cv}}^2 = 0.9547$). This model confirmed that molecular size-dependent bulk effects (dispersion and cavity formation) in the liquid state can be represented by a function of the gravitation index, whereas the hydrogen-bonding self-association interactions can be accounted for by the hydrogen donor charged surface area. However, the hydrogen-bonding donor structural features alone do not account for the differences among various hydrogen-bonding acceptors. These two descriptors need compensation by a topographic electronic index that corrects for the influence of hydrogen-bonding acceptors and branching effects in different isomers.

During recent years, neural networks have been applied to improve existing and develop new QSPR models for critical temperature. Hall and Story used electrotopological state indices to analyze the critical temperatures of a heterogeneous data set of 165 compounds.⁶⁰ The neural networks with 19:4:1 architecture gave a mean average error of 4.52 K for the whole set, 4.39 K for the training set, and 5.59 K for the test set. This model displays better performance from the point of view of errors in comparison with the MLR equations of Jurs⁶⁴ and Katritzky,⁶³ but includes 2 and 6 times more descriptors than these models. Indeed, considering the experimental error of 8 K, the data set seems to be overcorrelated.⁶⁴

Recently, Jurs and co-workers also used NN together with simulated annealing and genetic algorithms for descriptor selection to find accurate structure–property relationships for the same set of 165 compounds.⁶⁴ The 8:4:1 (41 adjustable parameters) architecture was used to train the network with 132 compounds, which resulted in a RMS error of 7.33 K. The prediction set with 18 compounds gave a RMS error of 9.85 K, which was reduced to 8.34 K by removing acetone (the biggest outlier) from the prediction set.⁶⁴

Vapor Pressure. Vapor pressure determines the volatility of a chemical compound. It governs the exchange rate of a chemical across an air–water interface through Henry's law constant. Accurate vapor pressures of chemicals of low volatility are often not available due to experimental difficulties. In such cases, the vapor pressure may be predicted using either the Clapeyron–Clausius equation and known values of the enthalpy of vaporization and the respective compressibility factor or the group contribution method. Quantitative structure–property relationships for the estimation of vapor pressures from descriptors derived solely from

the molecular structure offer a highly promising alternative which is more general and particularly suitable for the prediction of the vapor pressures of new chemical products. We have previously reviewed such QSPR analyses including those on small data sets⁶⁷ but concentrate here on those dealing with larger, diverse sets.

Chastrette et al. used a multifunctional autocorrelation method to predict the vapor pressure of alkenes and alcohols.⁶⁸ The components of the autocorrelation vector were calculated from Bondi's surface areas for the set of 186 compounds and used as descriptors in the analysis. The five-descriptor MLR equation with $R^2 = 0.988$, $s = 0.10$, was superior to previously published models.

Basak et al. used a hierarchical QSAR approach by relating the vapor pressures of 476 compounds from the ASTER database to topostructural, topochemical, and geometrical parameters.⁶⁹ The best MLR equation ($R^2 = 0.843$, $s = 0.29$) obtained consisted of 10 descriptors: 3 topostructural indices, 6 topochemical indices, and 1 hydrogen-bonding parameter. Examination of this correlation revealed that topological descriptors coding the chemical constitution and connectivity of atoms in the molecule play an important role in accounting for the variation in structures. However, inclusion of descriptors reflecting intermolecular interactions can improve the model substantially. In their final model, a simple hydrogen-bonding parameter shows the importance of H-bonding interactions in determining the vapor pressure of a compound. The relevance of descriptors related to intermolecular forces is obvious, because the more tightly the molecules are held together in a condensed phase, the lower the vapor pressure will be.

Liang and Gallagher employed a different set of 479 diverse compounds and quantum chemical descriptors to derive an MLR equation ($R^2 = 0.960$, $s = 0.534$)⁷⁰ dominated by the α -polarizability ($R^2 = 0.922$ for the one-parameter model) describing dispersion forces or induced dipole–induced dipole interactions. The correlation was improved by six functional group counts, covering dipole–dipole, dipole–induced dipole, and hydrogen-bonding interactions of polar compounds.

Our statistical analysis of an extensive set of constitutional, topological, electrostatic, geometrical, and quantum chemical descriptors led to a five-descriptor MLR equation for vapor pressure ($R^2 = 0.949$, $R_{\text{cv}}^2 = 0.947$, $s = 0.331$) for a set of 411 compounds.⁶⁷ This model indicates that vapor pressure is governed by structural factors similar to those already found for the boiling point³⁹ and critical temperature.⁶³ The gravitation index over all bonded atoms reflects the effective mass distribution in the molecule and effectively describes the molecular dispersion forces in the bulk liquid media. The hydrogen-bonding donor charged surface area also represents the forces of intermolecular attraction, particularly the hydrogen-bonding ability of the compound. Three additional descriptors compensate for an inadequate description of the intermolecular interactions occurring in molecules containing fluorine, chlorine, or nitrogen atoms. The cross-validated correlation coefficient shows the regression equation is of high stability and that the standard error approaches the experimental error of 0.32 log unit.

Flash Point and Autoignition Temperature. The flash point and autoignition temperature (AIT) are two closely related technological characteristics of flammability of

organic substances. The flash point is the temperature at which the vapor above a volatile liquid forms a combustible mixture with air. At the flash point, the application of a naked flame gives a momentary flash rather than sustained combustion, for which the temperature is too low. As the temperature rises still further, the combustible substances react with oxygen in the air in an exothermic oxidation process. Autoignition occurs when the rate of heat evolved by this reaction is greater than the rate at which heat is lost to the surroundings. The autoignition temperature is defined as the lowest temperature at which a substance in air will ignite in the absence of a spark or flame. The flash point and autoignition temperatures are important safety parameters in many situations related to the fire potential of chemicals. AIT is also crucial to the performance of internal combustion engines through the phenomenon of engine knock.

Correlations of flash points are rare and have given moderate results unless some provision has been made to separate compounds into similar functional groups. A modest correlation ($R^2 = 0.758$) was obtained for the flash points of 126 pyridines.⁵¹ A reduced data set of 121 pyridines with the exclusion of experimentally questionable data was used to develop a six-parameter equation for the flash points⁵² with $R^2 = 0.837$ ($R_{cv}^2 = 0.832$, $s = 16.7$ K). The descriptors employed in this equation indicate the importance of molecular bulk and hydrogen-bonding effects in determining the flash point.⁵²

In the first reported study of autoignition temperatures, Egolf and Jurs analyzed a data set of 312 hydrocarbons, alcohols, and esters.⁷¹ The general structure-based equation developed for the whole set was not very successful ($R^2 = 0.7259$). However, division of the set into four subsets [(i) low-temperature hydrocarbons, (ii) high-temperature hydrocarbons, (iii) alcohols, and (iv) esters] gave improved MLR equations especially for the low-temperature hydrocarbons. Recently, Mitchell and Jurs repeated this analysis with an extended set of descriptors. In addition new descriptor selection methods and a quasi-Newton BFGS algorithm based NNs were applied as well as the multilinear QSPR regression treatment.⁷² The attempt to develop a general MLR equation for the set of 300 compounds was still not satisfactory as it involved 11 descriptors and the RMS of the final equation was 58.5 °C. Better results were obtained by division of the structures into subsets of (i) low-temperature and (ii) high-temperature hydrocarbons, (iii) nitrogen-containing compounds, (iv) oxygen- and sulfur-containing compounds and (v) a subset of alcohols/esters. In most cases the NNs then gave improved results. The best models obtained were for subsets of high-temperature hydrocarbons ($R^2 = 0.8742$) and alcohols/ethers ($R^2 = 0.854$) with predictive qualities of RMS = 5.11 °C and RMS = 8.24 °C, respectively.

Tetteh et al. had used both radial basis and back-propagation neural networks to model autoignition temperature⁷³ with six descriptors, two of which were experimentally determined and the remaining four derived from the structure of compounds. A set of 85 compounds was used for training the networks, and AITs were then predicted for 148 compounds. The radial basis and back-propagation NNs led to satisfactory models for the training set ($R^2 = 0.9530$ and 0.9450) but performed much less well for the validation set ($R^2 = 0.8339$ and 0.8365 and average errors of prediction

of 30.1 and 29.9 °C, respectively). In a subsequent study, Tetteh et al. used biharmonic spline interpolation to optimize both the spread parameter and the number of neurons in the hidden layer of the radial basis function of NNs⁷⁴ in predicting the AIT of 232 organic compounds. This method gave moderate results for the training ($R^2 = 0.8259$, error 30.2 °C, $n = 78$), validation ($R^2 = 0.8334$, error 30.1 °C, $n = 77$), and test ($R^2 = 0.8612$, error 32.9 °C, $n = 77$) sets.

Density. The normal density (i.e., the density at 1 atm and 20 °C) is one of the major physicochemical properties used to characterize and identify a compound. It also provides an indication of its utility in certain industrial applications. In addition, densities can be used to predict or estimate other physical properties such as critical pressures.

A general QSPR treatment of 303 structures (containing C, H, N, O, S, F, Cl, Br, and I) incorporating a wide cross-section of classes of liquid organic compounds provided a good two-parameter correlation for densities ($R^2 = 0.9749$, $s = 0.0458$ for density ρ^{20}).⁷⁵ The main descriptor involved in this correlation represents the intrinsic density of the compound calculated as the ratio of the molecular mass and the molecular volume (represented by the overlapping van der Waals atomic spheres model) of the molecule. The second term is defined as the average electrostatic interaction per atom in the molecule, a term that is formally analogous to the Madelung energy in ionic crystals. Correlations were also developed for individual classes of organic liquids.

NNs have been applied to model densities of hydrocarbons by Gakh et al. with descriptors derived from graph theory. Training the NNs with 106 compounds resulted in an average error of 0.60% for a prediction set of 25 compounds.⁶²

Hu and co-workers studied the densities of alkenes in two parallel publications.^{59,58} They used 5 topological indexes for the set of 66 alkenes and obtained a correlation coefficient of $R^2 = 0.988$ with a relative standard deviation (RSD) of 0.61% for a prediction set of 16 compounds.⁵⁹ In the second study they used 5 slightly different topological descriptors and applied 5:3:1 NN architecture to train a set of 51 compounds, which resulted in an average RSD of 0.44% for the prediction set.⁵⁸

Refractive Index. The refractive index (η), defined as the ratio of the speed of light in a vacuum to the speed of light in the substance of interest, is one of the most important optical properties. It is frequently employed to characterize organic compounds and to evaluate the applicability of materials for various purposes. The refractive index bears a significant relationship to other optical, electrical, and magnetic properties, particularly the polarizability and dielectric permittivity of substances of high-field frequencies.

Although QSPR models have been previously developed for homologous series of compounds, a general QSPR relating the refractive index of a diverse set of 125 organic compounds with the chemical structure was proposed only recently in the form of a five-parameter correlation equation ($R^2 = 0.945$, $R_{cv}^2 = 0.937$, $s = 0.0155$).⁷⁶ The descriptors reflected the molecular polarizability, the charge distribution, hydrogen-bonding interactions, and molecular size dependency as underlying physical properties determining the value of the refractive index. For the external validation set of 25 randomly chosen diverse structures, the predicted values have an average error of 1.29%.

Neural networks have also been applied for the development of models for refractive indices. A back-propagation NN and 7 topological descriptors have been used to derive a model for the refractive indexes of hydrocarbons (106 in a training set and 25 in a prediction set) by Gakh et al.⁶² an average error of 0.16% was reached. Hu and co-workers have used 5 topological descriptors to predict the relative indexes of 51 alkenes with NNs having 5:5:1 architecture,⁵⁸ with an RSD of 0.11%. In a subsequent publication for a similar set of 66 alkenes, but using 5 different topological descriptors, they obtained a good correlation coefficient ($R^2 = 0.978$) with a 0.13% average RSD for a prediction set of 16 alkenes.⁵⁹

Melting Point. Melting point is another fundamental physical property of pure compounds. It specifies the transition temperature where the solid and liquid phases can coexist. Besides direct utility as an indicator of whether a compound is solid or liquid at given temperature, melting points have been applied in biochemical and environmental sciences due to their relationship with the solubility of compounds. Because of the complex interactions involved in the crystalline state, the melting temperature is expected to be a difficult property to describe by a uniform QSPR model for compound sets with large structural variability. The molecular packing in crystals is determined by molecular shape, size, and symmetry, hydrogen-bonding ability, and other intermolecular interactions such as charge-transfer and dipole-dipole interactions. All these interactions critically influence the melting point. Additionally, many compounds crystallize in more than one polymorphic form, with different melting points. Despite the vast amount of melting point data available and knowledge about the melting phase transition, no general relationship yet relates the melting point of compounds with their chemical structure. Most published quantitative structure-property relationships relating melting point to chemical structure are confined to limited and/or small sets of hydrocarbons, substituted aromatics, aldehydes, amines, and ketones (see ref 77 and references therein). We now review work on large sets.

Charton and Charton studied 366 congeneric alkanes and correlated both branched and unbranched compounds with an "intermolecular force equation"⁷⁸ which included a variable capable of accounting for the packing energy contribution of the alkyl group. The regression equation obtained with 11 descriptors had $R^2 = 0.9185$, with a standard error of 17.9 K. The contributions of the polar variables were slightly larger than those of the nonpolar variables, and structural variation in the substituent was more significant than that in the alkyl group.

The melting points of a diverse set of 443 mono- and disubstituted benzenes have been studied, resulting in a correlation equation with 9 descriptors ($R^2 = 0.8373$, $s = 30.19$ K).⁷⁷ Six-parameter equations for the subsets with *ortho*-, *meta*-, and *para*-disubstituted compounds were also presented. The importance of hydrogen-bonding descriptors is again reflected in these MLR equations. Notably, the same hydrogen-bonding descriptor (hydrogen donor charged surface area) was also important in the prediction of the boiling points.³⁹ Additionally, according to this model the melting point is influenced by molecular shape, size, and symmetry, related to the molecular packing in crystals, and by inter-

molecular interactions such as charge-transfer and dipole-dipole interactions.

Gramatica et al. have applied WHIM descriptors and genetic algorithms to predict the melting point of 209 polychlorinated biphenyls.⁷⁹ A test set of 82 compounds gave $R^2 = 0.82$ with 4 descriptors related to the size and symmetry of molecular structures.

III. INTERACTIONS BETWEEN DIFFERENT MOLECULAR SPECIES

Octanol-Water Partition Coefficient. The octanol-water partition coefficient is the most frequently used measure of hydrophobicity (or lipophilicity) of chemicals, which, in turn, is a very important property in medicinal chemistry, toxicology, and pharmaceutical and environmental sciences. Accordingly, there has been continuous significant interest in predicting octanol-water partition coefficients from chemical structure. As several excellent reviews are already available,^{80,81,82,83} we give only a brief overview.

The first method for estimating $\log P$ was pioneered by Hansch and Fujita, and based on the definition of the π -term for substituents.⁸⁴ However, most methods developed subsequently fall into three classes based on (i) fragment contributions, (ii) atomic contributions, or (iii) molecular descriptors.

(i) Fragment Contributions. An early calculation of partition coefficients from structure was developed by Rekker and his colleagues^{85,86} who derived 160 fragment contributions and several "proximity" correction factors using regression and statistical analysis from a data set of more than 1000 experimental $\log P$ values of simple organic compounds.⁸⁷ Their approach, labeled Σf , was later revised and extended.^{88,82} Rekker et al. noted that all the correction factors appeared to be multiples of the factor $C_M = 0.219$, which they attributed to hydration phenomena in functional groups.⁸⁸

CLOGP, perhaps the most extensively used method for the estimation of $\log P$ based on fragment constants was proposed by Hansch and Leo.^{89,90} In contrast to Rekker's "reductionistic" approach to derive fragment contributions by means of statistics, CLOGP's "constructionistic" approach assigns fragment values from the simplest possible molecules and applies numerous correction factors which account for fragment interactions.⁹⁰

(ii) Atomic Contributions. The first method for the calculation of $\log P$ as a sum of atomic contributions was proposed by Broto et al.,⁹¹ and included 222 atomic descriptors developed from 1868 chemical structures by regression analysis. These descriptors were derived from the classification of atoms according to their atom types and their topological environments. In a similar approach Ghose and Crippen^{92,93} used only 110 atomic contributions while maintaining the same standard deviation of prediction of 0.4 log unit. An extension by Viswanadhan et al. from 110 to 120 atom types derived from a data set of 893 compounds⁹⁴ was later extensively reevaluated with a much larger set of 8364 compounds.⁹⁵ A much larger set of 424 atomic and fragmental contributions was devised by Suzuki and Kudo.⁹⁶ Recently, Wang et al. calculated partition coefficients⁹⁷ for a data set of 1831 organic compounds. Atoms were classified by their hybridization state and neighboring atoms, and using

80 atomic and 5 correction factors, they obtained a standard deviation of 0.37 log unit. While atomic contribution approaches give good estimations in many instances and are easy to implement, they fail to account for long-range interactions, and an ever increasing number of atomic additives is required to maintain satisfactory predictions for increasing diversity.^{80,82}

Klopman et al. used the CASE program to identify group contributions and correction factors automatically for log *P* estimation.^{98,99} For 1663 diverse organic compounds they reported a correlation coefficient of $R^2 = 0.928$ and a standard deviation of 0.38 log unit while utilizing 94 parameters.⁹⁹ Meylan and Howard introduced the atom/fragment contribution (AFC) approach to estimate the partition coefficients of organic compounds.¹⁰⁰ By deriving 130 simple fragment contributions and 235 correction factors, they achieved a correlation of $R^2 = 0.98$ and a standard deviation of 0.22 for a training set of 2351 compounds and a correlation of $R^2 = 0.94$ and a standard deviation of 0.31 for the validation set of 6055 compounds.

(iii) Molecular Descriptors. Although the fragment contribution approaches give the most reliable predictions at present, they cannot estimate log *P* for structures containing a “new” fragment, and the number of correction factors continually grows as compound diversity increases. These shortcomings can be overcome by treating molecules as a whole rather than a sum of parts. Such an approach is taken by the third group of methods which predict an octanol–water partition coefficient from molecular properties or descriptors. Reflecting the “composite nature” of hydrophobicity/lipophilicity by encoding both steric and polar solvent–solute interactions, log *P* prediction methods have been proposed, which utilize such molecular properties as partial charge and charge densities, dipole moment, molecular surface area, volume, and electrostatic potential.⁸² Kamlet, Taft, and Abraham et al. developed an excellent correlation with a partition coefficient using empirical “solvatochromic” (or LSER) parameters for the solute’s volume, polarity/polarizability, and hydrogen-bonding acceptor/donor ability.^{101–103} Recently Luehrs et al. used LSER parameters to correlate partition coefficients of 981 diverse organic compounds with a standard deviation of 0.41.¹⁰⁴ However, the “solvatochromic” parameters have to be determined empirically, and their utility is inferior to theoretical descriptors, which can be derived directly from the molecular structure.

An original approach called TLSER (theoretical linear solvation energy descriptors) for the QSAR/QSPR to account for the solvent effects has been proposed by Famini.¹⁰⁵ The TLSER parameters have successfully been used for the correlation of various partition characteristics such as the adsorption on charcoal, Hafkensheid retention indices, and octanol/water partition coefficients.¹⁰⁶

The solvent-accessible surface area descriptors have been included in several predictive models for log *P*. Contributions to the surface area by the component atoms and group surface areas were used by Camilleri et al. to derive group contributions to partition coefficients for numerous benzene derivatives containing diverse substituents.¹⁰⁷ In a similar treatment, Masuda et al.¹⁰⁸ improved earlier results⁹³ by modifying atomic contributions to log *P* according to the degree of exposure of each atom to the surrounding solvent. This solvent-accessible surface area (SASA) scaled atomistic

method improved the correlation for the training set of 500 compounds from 0.869 for the original method to 0.948.

Politzer and co-workers combined molecular surface areas of the solute with electrostatic potential properties (average deviation and total variance of the electrostatic potential) to derive three- and four-parameter correlation equations with correlation coefficients (*R*) between 0.936 and 0.985 for a variety of organic compounds.^{109,66,110} The molecular geometries and electrostatic potentials were computed using the ab initio SCF method. Introduction of an “N factor”, the sum of the numbers of nitrogen and oxygen atoms in the molecule, improved the correlation.¹⁰⁹ Du and Arteca employed four surface electrostatic potential descriptors to correlate the partition coefficients of homologous series of aliphatic alcohols, amines, and acids, as well as of a set of functionalized aromatic compounds.¹¹¹ On the basis of ab initio calculations of electrostatic surface potentials (ESP), Sasaki et al.¹¹² used three ESP-derived parameters, surface tension, electrostatic interaction, and charge transfer, to correlate the partition coefficients of 63 solutes and obtained a regression coefficient of 0.983 and a standard deviation of 0.26 log unit.

Klopman et al. calculated the log *P* of simple organic compounds from semiempirical atomic charges and constitutional descriptors.^{113,114} Bodor et al. also employed semiempirical atomic charges along with other theoretical molecular descriptors including molecular volume, surface area, ovality, dipole moment, HOMO/LUMO energies, and atom counts to derive a nonlinear regression model for partition coefficients with correlation coefficient $R = 0.938$ and standard deviation 0.29 for a set of 118 organic compounds with 13 parameters.¹¹⁵ For an extended set of 302 compounds a QSPR model with 18 parameters had a correlation coefficient of $R^2 = 0.978$ and a standard deviation of 0.31.¹¹⁶ Kantola et al. calculated the conformationally dependent lipophilicity from charge densities, atomic contributions to the total molecular surface area, and dipole moments.¹¹⁷ The molecular geometries and partial charges were derived using semiempirical techniques. Molecular descriptors derived from semiempirical calculations were also used by our group to correlate the partition coefficients of 71 substituted pyridines.⁵² Using CODESSA descriptors, a six-parameter model with $R^2 = 0.943$ and standard deviation $s = 0.19$ was obtained which reflected the electrostatic and structural features of nitrogen atoms.

Topological and geometrical molecular descriptors have been also used to predict partition coefficients. Moriguchi et al. used atom-type descriptors together with factors for proximity effects, unsaturation, intramolecular hydrogen bonding, and ring structures to develop a 14-parameter regression equation which correlates the log *P* of 1230 solutes with a standard deviation of 0.41 and a regression coefficient of 0.952.^{118,119} Basak et al. used a stepwise regression analysis and a set of about 100 topological, topochemical, and geometric descriptors to derive an 11-parameter model with $R^2 = 0.91$ and $s = 0.35$ for 219 compounds.⁵⁰ Recently, Gombar and Enslein developed a robust model, based on a linear free energy relationship approach, for estimating log *P* on the basis of 363 electrotopological state indices and topological shape descriptors.¹²⁰ For the training set of 6675 diverse compounds they reported a correlation coefficient (R^2) of 0.986 and a standard deviation of 0.20; for a test set

of 47 nucleosides the standard deviation was 0.33.

To overcome the shortcomings of two-dimensional methods, Bodor and Buchwald had developed a 3D molecular size-based model to describe the lipophilicities of a variety of organic molecules using van der Waals molecular volume and a newly introduced parameter that should describe changes in hydrogen bonding during octanol–water transport.^{121,64} Their approach has been recently modified by Edward.¹²² A CoMFA-based three-dimensional technique to calculate the partition coefficient has also been reported by Waller.¹²³

Most of the models to calculate partition coefficients have been derived using regression analysis, and applications of neural networks are still rather limited. A combination of semiempirical solvent effect calculations and artificial neural networks enabled Grunenberg and Herges to predict the partition coefficients of a diverse set of organic compounds with a standard error of 0.29 log unit.¹²⁴ Eleven selected physical parameters from AM1 self-consistent reaction field calculations were used as input for neural networks of the back-propagation type. Among these 11 parameters, the solvation energy obtained by simple solvent calculations (spherical model) was found to be the most important parameter for predicting hydrophobicity.

Breindl et al. trained a back-propagation artificial neural net to estimate the log *P* values of a large range of organic molecules from the results of AM1 and PM3 semiempirical MO calculations.¹²⁵ The input descriptors included molecular descriptors such as electrostatic potentials, total dipole moments, mean polarizabilities, surfaces, volumes, and charges derived from semiempirical calculated gas-phase geometries. The estimates of log *P* for 105 organic compounds obtained by the resulting neural network had a standard deviation of 0.53 unit from the experimental log *P* values for AM1 and 0.67 unit in the case of PM3. A composite back-propagation neural network model (BNN) was developed by Devillers for calculating the log *P* of molecules containing nitrogen, oxygen, halogen; phosphorus, and/or sulfur atoms.¹²⁶ The molecules were described by the descriptors derived from components of an autocorrelation vector, **H**. The final 35:32:1 BNN predicted log *P* for an external test set of 519 molecules with correlation coefficient *R* = 0.98 and standard deviation *s* = 0.39.

Several reports compare the reliability of different methods to calculate log *P*.^{127,128,129,95,83} Some authors have found that the fragment contribution based methods deliver the most reliable predictions,¹²⁹ whereas others claim that the atom contribution technique is better for molecules with more than 45 atoms.⁹⁵

Aqueous Solubility of Liquids and Solids. Aqueous solubilities of liquids and solids are described by the parameter *S_w*, defined as the concentration (in units of moles or weight of solute per weight or volume of solution) of solute in the aqueous phase, at equilibrium with a pure solute phase. Access to the aqueous solubility of organic compounds is very important in many research areas, such as pharmaceutical or environmental science. It has been shown that aqueous solubilities are correlated with octanol–water partition coefficients.^{130,131,132,133,134,135,136}

Numerous different approaches to the prediction of aqueous solubility have been summarized by Yalkowsky and Banerjee¹³⁷ and classified into three categories: (i) correla-

tions with experimentally determined physicochemical quantities such as partition coefficient, chromatographic retention time, melting point, boiling point, molar volume (derived from liquid-density), or the parachor (derived from density and surface tension); (ii) correlations based on group contributions; (iii) correlations with parameters calculated solely from the molecular structure.

Yalkowsky et al. used the melting point and either the molecular surface area (MSA) or octanol–water partition coefficient for predicting *S_w* for polycyclic aromatic hydrocarbons¹³⁴ and halobenzenes.¹³⁸ The authors suggested that the melting point term accounts for the loss of lattice energy on solution of a solid. Dunnivant et al. also used melting point, MSA, and topological descriptors to predict *S_w* for polychlorinated biphenyls (PCBs).¹³⁹ Amidon et al. used MSA to describe *S_w* for a wide range of structures.¹⁴⁰ Taft and co-workers used the linear solvation energy relationship (LSER) approach for the description of *S_w* for aliphatic and aromatic hydrocarbons.¹⁴¹ Another promising theoretical approach to predict aqueous solubility¹⁴² combines the activity coefficient, γ (estimated from the water–octanol partition coefficient), with experimental parameters. Attempts to introduce corrections to account for the differences between the solubility of a liquid and a solid showed an improvement in the overall standard error. Isnard and Lambert¹³⁵ surveyed aqueous solubility values correlated to the 1-octanol–water partition coefficients in the literature and observed a 5–12% improvement in the standard error when the effects of crystalline interactions were accounted for by using the melting point of the solid compounds.

Group contribution methods for predicting solubility have been proposed by Klopman et al.,¹⁴³ Suzuki,¹⁴⁴ and Yalkowsky and co-workers.¹⁴⁵ Nirmalakhandan and Speece combined molecular connectivity indices with a modified polarizability and used it to predict aqueous solubilities of several sets of compounds,¹⁴⁶ but calculated their polarizability term by a different group contribution method for each set. A UNIFAC method was used by Li et al.¹⁴⁷ and by Kan and Tomson.¹⁴⁸ Although such models have attractive features, such as the ease of calculating solubility, they can neither explain the difference in solubility between isomers nor predict the solubility of a compound containing functionalities not present in the model.

Patil used molecular connectivity descriptors to predict aqueous solubility directly from molecular structure.^{149,150} Bodor and Huang used 18 descriptors, including semiempirically derived charge descriptors, to predict the solubilities of a set of 331 halogenated and oxygenated hydrocarbons with a standard error of only 0.299.¹⁵¹ Nelson and Jurs investigated a diverse set of 238 compounds,¹⁵² achieving a low prediction error by splitting the compounds into subsets according to their chemical classes (hydrocarbons, halogenated hydrocarbons, ethers, and alcohols) and using up to 9 descriptors. Sutter and Jurs also used 9 descriptors to achieve a standard error of 0.277 for another diverse set of 123 compounds.¹⁵³

The aqueous solubilities of a set of 96 hydrocarbons and 126 halogenated hydrocarbons (excluding compounds capable of forming hydrogen bonds) were correlated by a three-parameter equation involving geometrical, topological, and constitutional descriptors calculated solely from molecular structure, by Huibers and Katritzky.¹⁵⁴ The equation had a

correlation coefficient of 0.980 and a standard error of 0.386 log unit, compared to an estimated average experimental error of 0.24 log unit. The key descriptor was molecular volume, modified by topological and constitutional terms.

Correlation of the aqueous solubilities of 411 diverse organic compounds was reported by Katritzky and Karelson⁶⁷ using a general QSPR six-parameter correlation with correlation coefficient $R^2 = 0.879$, $R_{cv}^2 = 0.874$, and standard error $s = 0.573$, within the estimated experimental error of 0.58 log unit. The descriptors utilized were related to the polarizability of the molecule, cavity-size effects (dispersion and cavity formation), shape of the molecule, and specific solute–solvent interactions. These are the major determining factors for the solute–solvent interactions and, hence, aqueous solubility of compounds.

Many other authors had previously applied molecular descriptors and multiple linear regression to predict solubility. In addition, recent developments include the use of a mobile order solubility model,^{155,156} a clustering approach,¹⁵⁷ and neural networks. Bodor et al. used semiempirical descriptors to train neural networks to predict aqueous solubility.^{158,159} A 23:5:1 artificial neural network model reported by Huuskonen et al.^{160,161} related aqueous solubility values for 211 drugs and related compounds to topological and electrotopological indices and achieved predictive $R^2 = 0.86$ and $s = 0.53$ (log unit) for the test set. In another study, multiple linear regression (MLR) and computational neural networks (CNN) were utilized to develop mathematical models to relate the structures of a diverse set of 332 organic compounds to their aqueous solubilities.¹⁶² Topological, geometric, and electronic descriptors were used to numerically represent structural features of the data set compounds. Genetic algorithm and simulated annealing routines, in conjunction with MLR and CNN, were used to select subsets of descriptors that accurately relate to aqueous solubility. The result showed that nonlinear models with nine calculated structural descriptors were developed that have a training set root mean square error of 0.394 log unit for compounds which span a $-\log(\text{molarity})$ range from -2 to $+12$ log units.

Aqueous Solubility of Gases and Vapors (Water–Air Partition Coefficients). Due to the technical difficulties of an accurate analytical determination of the solubility of gases and vapors, methods for their accurate prediction are of great practical importance. The solubility of a gas or vapor in water is usually described by the water–air partition coefficient (L_w), also known as the Ostwald solubility coefficient, which is defined as the ratio of the concentration of a compound in an aqueous solution and in the gas phase at equilibrium. Another commonly used parameter is Henry's law constant H , which is essentially an air–water partition coefficient and is approximately equal to L_w^{-1} . Water–air partition coefficients can be estimated from the vapor pressure (VP) and the aqueous solubility (S_w) of a compound^{163,164,165} using a thermodynamic cycle or directly from the molecular structure on the basis of a quantitative structure–property relationship (QSPR).

Hine and Mookerjee probably reported the first empirically based group and bond contribution schemes.¹⁶⁶ Their group contribution scheme reproduced the solubilities of 292 diverse compounds with a standard error of 0.12 log unit but used 69 empirical group contribution factors. Their bond contribution scheme reproduced the solubilities of 263

diverse solutes with a standard error of 0.42 log unit using 34 bond contributions. Another group contribution scheme was developed by Cabani et al.¹⁶⁷ who used 28 group contributions to reproduce 209 log L_w values of diverse compounds to within 0.09 log unit. Because of the large number of fitted parameters involved in these schemes, neither the group contribution nor the bond contribution method conveys much understanding of the physical nature of the relationship between the molecular structure and interactions and the solubility of gases in water. Moreover, their application to the prediction of solubilities of compounds containing structural functionality not included in the original set is not justified.

Nirmalakhandan and Speece¹⁶⁸ developed a predictive model involving three structurally determined descriptors: the valence connectivity index,⁸ a molecular polarizability descriptor, Φ ,¹⁶⁹ and an indicator variable, I , for the presence of an electronegative atom. In their study, the solubility data for 180 diverse compounds were reproduced with a standard error of 0.262 log unit. However, the polarizability descriptor was calculated on the basis of an atomic contribution scheme involving another 11 empirical parameters, which effectively increases the number of parameters employed to 14.

Russell, Dixon, and Jurs correlated the logarithms of Henry's law constant, $\log H$, of a limited data set of 63 diverse gases in water, using 5 theoretically calculated descriptors.¹⁷⁰ Their linear regression model had a correlation coefficient of 0.978, a standard deviation of 0.375 log unit, and an F -statistic of 250. The authors suggested that the factors influencing the solubility of gases in water were related to the solute bulk, lipophilicity, and polarizability.

Abraham et al. correlated the solubility of 408 diverse gases in water with 5 linear solvation energy relationship (LSER) descriptors¹⁶⁴ including the excess molar refraction (calculated from the experimental molar refraction), the experimentally determined dipolarity/polarizability π_2^H , the effective hydrogen-bond acidity $\Sigma\alpha_2^H$ and basicity $\Sigma\beta_2^H$, and the McGowan characteristic volume V_x (calculated from some tabulated atomic increments). The model had a correlation coefficient of 0.9976, a standard deviation of 0.151 log unit, and an F -value of 16 810. Although this correlation equation can be interpreted term-by-term using well-established chemical principles, the LSER descriptors cannot be used to make a priori predictions because four of these descriptors have to be determined experimentally. Also, since the resulting correlation does not relate the property to the molecular structural information, it is difficult to elucidate how molecular structure affects the observed property.

The partitioning of two sets of organic gases and vapors between water and air (L_w) has been studied using the CODESSA program.¹⁷¹ For the first set of 95 alkanes, cycloalkanes, alkylarenes, and alkynes, excellent predictions were obtained with a two-parameter correlation equation ($R^2 = 0.977$, $R_{cv}^2 = 0.975$, $s = 0.20$). The two descriptors involved, the gravitation index and the complementary information content, reflect the effective mass distribution and the degree of branching of the hydrocarbon molecule, and adequately represent the effective dispersion and cavity formation effects for the solvation of nonpolar solutes in water. For the second set of 406 widely diverse organic compounds (including structures containing N, O, S, and

halogen atoms), a successful 5-parameter correlation equation ($R^2 = 0.941$, $R_{cv}^2 = 0.939$, $s = 0.53$) was also reported. The descriptors from this equation (which were completely different from those for the set of 95 nonpolar solutes) comprised the partial charge weighted normalized hydrogen-bonding donor surface area, counts of oxygen and nitrogen atoms, the HOMO–LUMO energy gap, the most negative partial charge weighted topological electronic index, and the number of rings. They account for the dispersion energy of polar solutes in solution, the electrostatic part of the solute–solvent interaction, and hydrogen-bonding interactions in liquids.

In a related study, water–air partition coefficients were estimated from the vapor pressure and aqueous solubility, which, in turn, were predicted by the QSPR models derived from a diverse set of 411 compounds.⁶⁷ The mean standard error of the predicted gas solubilities was found to be very similar to the standard error of the L_w , predicted using the equation derived directly from the experimental values of L_w .¹⁷¹

Solvent Polarity Scales. The use of solvents is fundamental to the practice of chemistry, and the choice of an appropriate solvent is anything but trivial. To assist chemists in their understanding of solvent properties and in the choice of solvent, many solvent polarity scales have been developed.¹⁷² These scales are based on diverse physicochemical phenomena including reaction rates, solvatochromic effects, reaction enthalpies, etc.^{173,174,175,176} Frequently the actual mechanism of the influence of the solvent on a physical or chemical process is unclear. The same is often true about the individual polarity scales.

A three-parameter MLR equation with $R^2 = 0.936$ ($R_{cv}^2 = 0.900$) was developed for the unified nonspecific solvent polarity scale (S') on the basis of theoretical molecular descriptors.¹⁷⁷ It correlates S' for 25 structurally diverse solvents within a 5% average absolute error. The correlation equation includes the following three orthogonal theoretical molecular descriptors: (i) the average structural information content (zeroth order), (ii) the weighted partial negative surface area, and (iii) the hydrogen-bonding acceptor surface area. These descriptors provide insight into nonspecific solvation at the molecular level. They reflect adequately the solvent–solute interactions in the internal cavity of the solvents. Predictions using this three-parameter model are used to extend available S' values to a total of 67 solvents.

The same solvent polarity scale has also been studied using model-based QSPR (MQSPR) to enable the prediction of the S' values from quantum-mechanical calculations.¹⁷⁸ MQSPR selects descriptors prior to the correlation analysis that have potential meaning in the content of the measured property. A successful two-parameter correlation ($R^2 = 0.9587$) is developed from 48 diverse molecules using two orthogonal descriptors, the dipolar density (given by the total dipole moment of the molecule divided by the molecular volume) and the reciprocal of the HOMO–LUMO energy gap. The correlation equation allows confident estimation of S' from quantum-mechanical calculations, assists in the interpretation of nonspecific solvation, and enables one to analyze solvent conformational-dependent solvation influences.

In a more comprehensive study, the most important solvent polarity scales were collected and QSPR models developed

for each of them. Altogether 45 different solvent polarity scales and 350 solvents were analyzed. QSPR models for each of the 45 scales were constructed using only theoretical descriptors: of the 45 models, 27 give $R^2 > 0.90$ and only 2 had $R^2 < 0.82$.¹⁷⁹ This study allowed a unified PCA treatment of solvent polarity where the missing values in the polarity scales are calculated from correlation models derived using the CODESSA program.¹⁸⁰ A set of 40 scales and 40 solvents showed that the 3 main principle components accounted for 74% of the total variance. Moreover for 27 of the 40 scales, these 3 components described $\geq 85\%$ of the variance and for only 4 of the scales was less than 70% of the variance described. The PCA loadings showed clear clustering of the scales in a three-dimensional space in a chemically rational manner. Similarly, the PCA scores classified the solvents intelligently.¹⁸⁰

GC Retention Time and Response Factor. Due to the high popularity of gas chromatography as an analytical technique, there has been a significant interest in developing methods for predicting chromatographic retention times and retention indices.¹⁸¹ An excellent overview of the theoretical basis of the quantitative structure–retention relationships is given by Kaliszan.²³ Most reported correlations between the molecular structure and GC retention indices either relate to narrow classes of compounds^{182,183,184,185,186,187,188,189,190,191,52,192,193} or utilize other experimentally-determined physicochemical properties.¹⁹⁴

Methods that can predict chromatographic retention times solely from the molecular structure are obviously of greater utility. Several authors have discussed the estimation of retention indices using topological descriptors.^{195,196,197} Duvénbeck and Zinn reported a general method for fitting the gas chromatographic retention index data using three topological indices and one electrotopological state index in the so-called vertex and edge multilinear regression models.¹⁹⁸ For a data set of 217 acyclic and cyclic alkanes, alkenes, alcohols, esters, ketones, and ethers, their edge model gave a mean absolute error of 9 retention index units. However, application of this model for the prediction of the retention indices for test compounds of the same classes gave prediction errors from 15 to 22 retention index units.¹⁹⁹

Stanton and Jurs introduced the charged partial surface area (CPSA) descriptors,⁴⁵ which were used in combination with the topological and geometrical descriptors to predict retention indices of substituted pyrazines,²⁰⁰ polycyclic aromatic compounds,¹⁸³ stimulants and narcotics,²⁰¹ and anabolic steroids.²⁰² Descriptors in these multilinear regression analysis (MLR) models encode information related to the types of interactions that take place between the solute (and solvent) molecules in the stationary phase during the separation process. As the polarity of the stationary phase is changed, different descriptors become important, and therefore, each phase has to be modeled separately.

Buydens, Massart, and Geerlings combined topological descriptors with quantum-chemical descriptors to predict the GC retention indices of mono- and bifunctional alcohols and ketones.²⁰³ In a more general QSPR study, a mixed set of topological and quantum-chemical descriptors was used to correlate GC retention times of 152 structures from a wide cross-section of classes of organic compounds.²⁰⁴ A forward procedure for the selection of molecular descriptors for multilinear regression analysis in the CODESSA program

gave a six-parameter model ($R^2 = 0.959$, $R_{cv}^2 = 0.955$, $s = 0.515$), in which the AM1-computed α -polarizability was the most important descriptor. These results were recently reevaluated using improved procedures in CODESSA and new methods for the efficient selection of variables in the multilinear regression analysis.²⁰⁵

Quantum-chemical descriptors were also employed by Donovan and Famini in a theoretical linear solvation energy relationship (TLSER) investigation of the GC retention indices of 37 organosulfur compounds.²⁰⁶ From each of the three semiempirical methods (MNDO, AM1, and PM3) used to compute six TLSER descriptors, similar correlations were obtained: $R = 0.94$ – 0.96 . These results were of quality statistically similar to that of the previous study of the same compounds by Woloszyn and Jurs.²⁰⁷ However, the TLSER approach²⁰⁶ was also able to handle compounds containing sulfur–sulfur bonds, which were omitted in the four parameter correlation with topological and CPSA descriptors obtained by Woloszyn and Jurs.²⁰⁷

Whereas the majority of the structure–retention relationship predictions of GC retention indices are based on multilinear regressions, recent significant studies have employed neural networks. Bruchmann, Zinn, and Haffer showed that neural networks can be trained using electrotopological indices of monofunctional compounds by the back-propagation technique to predict the corresponding retention index data.²⁰⁸ They used 29 input neurons, two hidden neurons, and one output neuron in the respective NN. However, it has to be noted that whereas the predictions by the trained neural network were similar to those by a multilinear regression model for the training set, they were much poorer in the case of extrapolating predictions for the test set of terpene compounds.

Sutter, Peterson, and Jurs applied neural networks to predict retention indices of alkylbenzenes from their molecular structure.²⁰⁹ They used ADAPT software to calculate 182 descriptors and a multilinear regression analysis in combination with evolutionary optimization algorithms to select a subset of descriptors relevant to mapping retention indices. Six descriptors from their best multiple regression model (three topological, one geometrical, and two CPSA descriptors) were used with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method to train a 6:5:1 neural network which improved RMS for both training and prediction sets from 18.0 and 21.8 to 13.4 and 17.6, respectively.

Recently, an application using counterpropagation neural networks for the prediction of GC retention indices was reported.²¹⁰ The counterpropagation NN is composed of two layers: the Kohonen layer which acts as a single-layer Kohonen network²¹¹ and an output layer. A 15 by 15 counterpropagation NN was trained by presenting 12-dimensional vectors of eight topological and four constitutional descriptors. For a test set, GC retention indices predicted by the trained NN correlated well with experimental values ($R^2 = 0.92$). However, this correlation was slightly worse than that reported in the previous study of the same data set with back-propagation NNs.²⁰⁸ In a follow-up study using an extended data set of 381 simple organic compounds, it was found that predictions by the counterpropagation NN model were inferior to those by the back-propagation NN or multilinear regression models.²¹² In fact, the 16:12:1 back-propagation NN model gave the lowest

RMS error in a 10-fold cross-validation procedure when compared to the 16-parameter MLR model and a 25×25 counterpropagation NN model (19.2, 22.5, and 36.1, respectively). A set of 16 topological and constitutional descriptors was used to fit all models.

Flame ionization detector (FID) response factors are used in gas chromatography to measure amounts of the compounds analyzed. The prediction of such response factors is of particular significance since they are independent of the GC column parameters. Scanlon and Willis estimated response factors on the basis of the “effective carbon number” descriptor.²¹³ Our groups studied 152 diverse compounds and reported a good six-parameter MLR ($R^2 = 0.892$, $R_{cv}^2 = 0.881$, $s = 0.0543$) which can be used to predict response factors for many classes of organic compounds.²⁰⁴ The most important descriptor was found to be the relative weight of the “effective” carbon atoms. These results were recently reevaluated using the improved procedures in CODESSA and new methods for the efficient selection of variables for the multilinear regression analysis.²⁰⁵ The same data set has also recently been analyzed by Jelali-Heravi and Fatemi, who used both multilinear regression and traditional back-propagation NN techniques²¹⁴ and reported similar performance by both MLR and NN models (SE 8.95 vs 6.57 for the training set, and 5.37 vs 7.12 for the prediction set, respectively).

IV. SURFACTANT PROPERTIES

Critical Micelle Concentration. The critical micelle concentration (cmc) is an extremely useful quantity, which describes the ability of dissolved surfactants to reduce surface or interfacial tension. A low cmc indicates that it is thermodynamically favorable for the hydrophobic domain of the surfactant molecule to leave the aqueous solution, and this results in both an excess concentration at the interface and the formation of micelles. This ability to absorb at an interface and reduce interfacial tension is of great importance to many processes of technological interest, such as emulsification, foaming, wetting, solubilization, detergency, particle suspensions, and surface coatings. The cmc is also important in thermodynamic studies of micellar solutions, inasmuch as it is related to the free energy of micelle formation. Before our work several QSPR analyses of homologous series of surfactants had appeared; see refs 215 and 216. We have established more general structure–cmc relationships through a systematic quantitative structure–property relationship approach, using the large database of molecular descriptors and the strategies implemented in CODESSA. For these studies, fragment descriptors, based on the two clearly differentiated hydrophobic and hydrophilic fragments of a surfactant, as well as those for the whole molecule, were tested.

The QSPR study of the critical micelle concentrations of a diverse set of 77 *nonionic surfactants*²¹⁵ produced a general three-parameter structure–property relationship ($R^2 = 0.9833$, $R_{cv}^2 = 0.9812$, $s = 0.1769$). Two of the three descriptors represent contributions from the topology of the hydrophobic group, and the third is a measure of the size of the hydrophilic group. The cmc of nonionic surfactants in aqueous solution is primarily determined by the hydrophobic part of the molecule. The logarithm of the cmc decreases with an

increase in the size of the hydrophobic fragment and increases with an increase in the relative size of the hydrophilic fragment. Hydrophobicity is affected by the branching of the hydrophobic fragment and by the presence of heteroatoms.

QSPR investigation of relationships between the molecular structure and the measured cmc for 119 *anionic surfactants* comprising sodium alkyl sulfates and sodium sulfonates with a wide variety of hydrophobic tails²¹⁶ gave an optimum three-descriptor, multiple linear regression model with a correlation coefficient of $R^2 = 0.940$ ($s = 0.2173$). The correlation equation contains qualitative contributions from both the hydrophobic and the hydrophilic domains of the surfactant molecules, as well as the dipole moment representing a descriptor for the entire molecule. The nature of the selected descriptors is in agreement with known qualitative structural effects on cmc. A still better correlation ($R^2 = 0.986$) was obtained using 3 descriptors for a subset of 63 structures, with variation only in the hydrophobic domain.

Cloud Point. The cloud point is an important property of nonionic surfactants. Below this temperature a single phase of molecular or micellar solution exists; above it the surfactant has reduced water solubility, and a cloudy dispersion results. A general empirical relationship ($R^2 = 0.937$, $s = 6.5$) has been developed for estimating the cloud point of pure nonionic surfactants of the alkyl ethoxylates using only topological descriptors.²¹⁷ The set of 62 structures is composed of linear alkyl, branched alkyl, cyclic alkyl, and alkylphenyl ethoxylates. For this set the cloud points can be estimated to an accuracy of ± 6.3 °C (3.7 °C median error) using the logarithm of the number of ethylene oxide residues and three topological descriptors that account for hydrophobic domain variation. The topological descriptors model various aspects of the hydrophobic tail structure.

Bünz et al. also used theoretically derived descriptors together with MLR analysis and feed forward NN to derive structure–property-based relationships for surfactants.⁴² Cloud point data of 23 alkyl ethoxylate surfactant molecules provide four descriptor MLR equations with $R^2 = 0.9948$ and $s = 1.66$ K which combined two topological and two constitutional descriptors.⁴²

V. COMPLEX PROPERTIES AND PROPERTIES OF POLYMERS

Polymer Glass Transition Temperature. The QSPR description of polymer properties represents an interesting challenge, since many theoretical molecular descriptors for high molecular weight compounds are difficult to calculate, or cannot be calculated directly. The glass transition temperature, T_g , also known as the glass temperature or the glass–rubber transition temperature, is one of the most important properties of amorphous polymers. In the vicinity of T_g , a polymer experiences a sudden increase in the rate of molecular motions and, as a result, undergoes a series of conformational transformations. The torsion oscillations and/or rotations about most of the backbone bonds are activated, which causes a sharp increase in the free volume of the system as it is converted from the initial rigid (glassy) state to the quasi-liquid state. The most complete recent publication about various methods for prediction of polymer properties, including QSPR, is compiled by Bicerano.²¹⁸ An

overview of QSPR-related methods and models derived is also provided in our previous publications.^{219,220}

Camelio et al. correlated the T_g values of 50 acrylates and methacrylates with their “backbone flexibility” as represented by the ratio of the total energy of a repeating unit inside the cylinder relative to the volume of this cylinder.²²¹ All necessary parameters were calculated using molecular mechanics. This method resulted in a correlation with an R^2 of 0.83.

Rode and Tam collected a set of 62 poly(acrylic acid), poly(methacrylic acid), and polyacrylamides to analyze the applicability of Mulliken charges from AM1 and PM3 parametrizations and Gasteiger–Huckel charges for the construction of QSPR models in predicting glass transition temperatures.²²² For the final analysis, the six most important atoms in the monomer structure were considered along with the degree of substitution (m) in the monomers and the chain length (n) of the hydrocarbon group. In this way, eight descriptors were involved in the final correlation model. The importance of the net charge in the polymer may be understood by its role in cohesive force. The permanent changes in molecules cause a permanent dipole that interacts with neighboring dipoles and produces a polar cohesive force. Therefore, any changes in the charges will lead to changes in T_g . The statistical characteristics clearly show the better performance of PM3 charges and especially AM1 charges compared to the Gasteiger–Huckel charges.

Waegell and co-workers studied a small set of acrylate and methacrylate polymers with the “EVM model” for QSPR analysis (with $R^2 = 0.96$ and $s = 12$ K).^{223,224} EVM stands for the three descriptors, energy, volume, and mass, involved in the QSPR model. The energy was related to the polymer conformations and was estimated using a molecular mechanics method. The volume descriptor coded the space occupied by the atoms as well as the space available for their movements. The mass term was the molecular weight of the monomer or repeating unit. For linear and branched aliphatic acrylate and methacrylate polymers, the standard deviation from linear regression was 12 K, and the $R^2 = 0.96$ for 16 compounds in the training set. The model allows the calculation of the T_g with an average absolute error of 10% for 34 linear and branched derivatives not included in the original QSPR model.

A preliminary successful QSPR analysis of 22 linear, medium molecular weight polymers and copolymers²¹⁹ encouraged us to develop a more general model for the molar glass transition temperature (T_g/M). An analysis of a diverse set of 88 high molecular weight polymers resulted in a five-parameter QSPR model ($R^2 = 0.946$ and the standard error 0.33 K g mol⁻¹).²²⁰ The descriptors were calculated for the middle unit of the trimeric structure of three repeating units for each polymer; in this way the influence from adjacent repeating units was taken into account. The descriptors in the model relate to the rotational flexibility of the molecules at the T_g , the branching of the polymer molecules, hydrogen-bonding interactions, and electrostatic interactions between the polymer molecules. This approach is applicable, in principle, to all linear polymers of regular structure, and encourages the further application of QSPR methods to other types of polymers such as copolymers, cross-linked polymers, and biopolymers.

Rogers and Hopfinger used the genetic function approximation (GFA), an extension of the genetic algorithm, to obtain QSPR models for polymer glass transition temperatures.²²⁵ A dataset of 35 compounds resulted in a six-parameter QSPR model with $R^2 = 0.910$. The descriptors involved were the backbone and side chain contributions to the monomer conformational entropy, backbone mass moments, and positive and negative electrostatic intermolecular energies for the complete monomer unit. Luke applied evolutionary programming to the same set of compounds, to compare the EP with the GFA approach and test the applicability of the EP method in developing QSPR models;²²⁶ he concluded that the two methods were equal in quality, but that EP found several good QSPR models missed by the GFA.

Noid and co-workers introduced PropNet technology which utilizes neural networks as a tool to derive structure–property relationships.²²⁷ Up to 320 different polymers (T_g ranging from 50 to 700 K, including tactic and cross-linked polymers) were used to test TgNet (the individual network for prediction of the polymer glass transition temperature) for robustness and accuracy. The results demonstrate that TgNet is capable of predicting the glass transition temperature to within 10 K of the experimentally reported values.

Joyce et al. used monomer structures alone to predict physical properties of polymers with back-propagation neural networks.²²⁸ The networks were trained using 360 example monomer structures and tested on an independent set of 89 different monomer structures. The best networks were able to predict the T_g values for a testing set of polymers with a wide range of structures with a RMS error of about 35 K.

Polymer Refractive Index. The refractive index (η) of a polymer is a fundamental optical property directly related to other optical, electrical, and magnetic properties and widely used in material science as an indication of the potential of a polymer for specific purposes. A satisfactory quantitative structure–property relationship allowing quantitative predictions of the refractive index of both known and of as yet unsynthesized polymers would clearly be of significant utility.

Bicerano²¹⁸ developed a QSPR relationship ($R^2 = 0.955$) to predict refractive indices for a set of 183 polymers with 10 descriptors. These descriptors included three different topological indices, the total number of rotational degrees of freedom (both of the polymer backbone and of the side groups), and several constitutional descriptors such as the number of chlorine atoms, the number of chlorine atoms bonded to an aromatic ring, the number of sulfur atoms, and the number of hydrogen-bonding moieties.

A general MLR equation ($R^2 = 0.940$, $R_{cv}^2 = 0.934$, $s = 0.018$) was developed using CODESSA for the correlation of the refractive indices for a diverse set of 95 amorphous homopolymers.²²⁹ The five descriptors involved in the model are calculated from the structure of the repeating unit of the polymer. The final equation was derived with an intercept fixed to a value of one, *i.e.*, to the refractive index of a vacuum. The correlation model shows that the polarizability (described by the HOMO–LUMO energy gap) has an important influence on the refractive index of polymers just as for low molecular weight compounds (see above).²²⁹ Compounds of lower stability (as described by the heat of formation) possess higher refractive indices. Other descriptors

show the importance of charge distribution and the hybridization of carbon atoms in the repeating unit of the polymer. The average prediction error of the model is 0.9%, and the highest prediction error is 3.2%.

Rubber Vulcanization Acceleration. In spite of the fact that the vulcanization of rubber has been studied for many years, its precise mechanism has remained unclear. A CODESSA QSPR treatment has assisted in the understanding of some key features of this process. Regression analysis correlated various parameters, including *ts2* (the onset of cure) and *mxr* (the maximum rate of vulcanization), with molecular descriptors.²³⁰ Two classes of accelerator molecules studied comprised (i) disulfides and (ii) a combination of sulfenamides and sulfenimides. Parent molecules of the accelerators and also their zinc complexes with thiolate fragments were modeled for each class to give a total of four data sets, all of which gave good correlations.

VI. CONCLUSIONS

The results reviewed in this paper witness the applicability and power of the quantitative structure–property relationship approach for the prediction of very diverse technology-related properties of chemical compounds and materials. This has become possible due to substantial progress in the development of new, more adequate molecular descriptors and methods of derivation of multiple linear and nonlinear relationships. In many cases, the QSPRs not only represent empirical equations for formal interpolation or extrapolation of missing data but give insight into the physical interactions and processes determining the properties of substances. Moreover, the ability to use exclusively theoretical molecular descriptors has provided the means to predict the properties for molecular structures that are difficult to measure experimentally or even for those not yet synthesized.

The QSPR methodology being developed forms a solid basis for the creation of a genuine chemical artificial intelligence. Supported by advancements in computer hardware and software technology, it is now realistic for a generation of computational chemists to tackle this scientific challenge. The use of massive parallel processing, together with the QSPR and quantum chemistry-based data mining, and case-based and rule-based reasoning technologies, should enable an effective approach to the so-called reverse problem—the reliable prediction of chemical structures with predetermined chemical, physical, and technical properties. Undoubtedly, such a computerized chemical mind would become a most powerful theoretical tool for all industrial, environmental, or pharmaceutical chemists.

Future directions in QSPR methodology evolve from the need to predict the properties of complex multicomponent and high molecular weight molecular systems. Specific molecular descriptors will have to be developed for the accurate and reliable characterization of such systems. At elevated temperatures and pressures, many molecules do not have a fixed geometrical structure but exist as a Boltzmann distribution over a variety of molecular conformations and configurations. The formulation of QSPR equations adequate for such conditions is anything but trivial. A possible solution to this question may lay in the appropriate combination of QSPR with the statistical mechanics and quantum chemistry. Recent advances in these fields give optimism concerning the development of the methodologies required.

In the real world, most systems are multicomponent, frequently with mixture of isomers and homologs. The application of QSPR to mixtures will need to be tackled.

Again, in many important areas such as toxicity, there are numerous alternative quantitative measures of a property. We need to develop methods to prove relationships between such characteristics as human, mammalian, insect, fish, and bacterial toxicity, and between acute and chronic toxicity. The combination of QSPR with PCA, as used in the study of solvent polarity, may show the way to greater insight into such problems.

In conclusion, we hope that the present paper may give an additional impulse to computational chemists to address the *real-life* problems and challenges in chemical technology and materials science.

REFERENCES AND NOTES

- Martin, Y. C. 3D QSAR: Current State, Scope, and Limitations. *Perspect. Drug Discov.* **1998**, 12, 3–23.
- Norinder, U. Recent Progress in CoMFA Methodology and Related Techniques. *Perspect. Drug Discov.* **1998**, 12, 25–39.
- Maddalena, D. J. Applications of Soft Computing in Drug Design. *Expert Opin. Ther. Pat.* **1998**, 8, 249–258.
- Kubinyi, H. QSAR and 3D QSAR in Drug Design. 2. Applications and Problems. *Drug Discov. Today* **1997**, 2, 538–546.
- Hansch, C.; Fujita, T. Status of QSAR at the End of the Twentieth Century. In *Classical and Three-Dimensional QSAR in a Agrochemistry*; Hansch, C., Fujita, T., Eds.; American Chemical Society: Washington, DC, 1995; 606, pp 1–12.
- Hansch, C.; Leo, A. *Exploring QSAR, Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- Stuper, A. J.; Brugger, W. E.; and Jurs, P. C. *Computer-assisted Studies of Chemical Structure and Biological Function*; While: New York, 1979.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; J. Wiley & Sons: New York, 1986.
- Murray, J. S.; Politzer, P. A General Interaction Properties Function (GIPF): An Approach to Understanding and Predicting Molecular Interactions. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 243–289.
- Hilal, S. H.; Carreira, L. A.; Karickhoff, S. W. Estimation of Chemical Reactivity Parameter and Physical Properties of Organic Molecules Using SPARC. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 291–353.
- Abraham, M. H. New Solute Descriptors for Linear Free Energy Relationships and Quantitative Structure-Activity Relationships, In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 83–133.
- Abraham, M. H.; Chadha, H. S.; Dixon, J. P.; Rafols, C.; Treiner, C. Hydrogen Bonding. Part 40. Factors that Influence the Distribution of Solutes Between Water and Sodium Dodecylsulfate Micelles. *J. Chem. Soc., Perkin Trans. 2* **1995**, 887–894.
- Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279–287.
- Randic, M.; Razingar M. On the Characterization of Three-Dimensional Molecular Structure. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum Press: New York, 1996; pp 159–236.
- Balaban, A. T. From Chemical Topology to 3D Geometry. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 645–650.
- Lucic, B.; Trinajstić, N. Multivariate Regression Outperforms Several Robust Architectures of Neural Networks in QSAR Modelling. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 121–132.
- Jurs, P. C.; Chou, J. T.; Yuan, M. Studies of chemical structure-biological activity relations using pattern recognition. In *Computer-Assisted Drug Design*; American Chemical Society: Washington, DC, 1978; Vol. 112, pp 103–129.
- Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. QSPR and QSAR Models Derived Using Large Molecular Descriptor Spaces. *Collect. Czech. Chem. Commun.* **1999**, 64, 1551–1571.
- Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, 96, 1027–1043.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983.
- Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structure*; Wiley-Interscience: New York, 1983.
- Kalisan, R. *Quantitative Structure-Chromatographic Retention Relationships*; J. Wiley & Sons: New York, 1987.
- Merrifield, R. E.; Simmons, H. E. *Topological Methods in Chemistry*; J. Wiley & Sons: New York, 1989.
- Rouvray, D. H., Ed. *Computational Chemical Graph Theory*; Nova Science Publishers: New York, 1990.
- Myers, R. H. *Classical and Modern Regression with Applications*; PWS-KENT Publishing Co.: Boston, 1989.
- Draper, N. R.; Smith, H. *Applied Regression Analysis*; J. Wiley & Sons: New York, 1981.
- Malinowski, E. R.; Howery, D. G. *Factor Analysis in Chemistry*; Wiley-Interscience: New York, 1980.
- Strouf, O. *Chemical Pattern Recognition*; J. Wiley & Sons: New York, 1986.
- Meloun, M.; Militky, M.; Forina, M. *Chemometrics in Analytical Chemistry*; Ellis Horwood: New York, 1992.
- Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH Publishers: Weinheim, 1993.
- Mitchell, T. M. *Machine Learning*; McGraw-Hill: New York, 1997.
- Gini, G.; Testaguzza, V.; Benfenati, E.; Todeschini, R. Hybrid Toxicology Expert Systems: Architecture and Implementation of a Multi-domain Hybrid Expert Systems for Toxicology. *Chemometr. Intell. Lab.* **1998**, 43, 135–145.
- Langmuir, I. The Distribution and Orientation of Molecules. *Colloid Symp. Monogr.* **1925**, 3, 49–75.
- Hermann, R. B. Theory of Hydrophobic Bonding. II. The Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area. *J. Phys. Chem.* **1972**, 76, 2754–2759.
- Reichsteiner, C. E. In *Handbook of Chemical Property Estimation Methods*; Lyman, W. J., Reehl, W. F., Rosenblatt, D. H., Eds.; McGraw-Hill: New York, 1982; Chapter 12.
- Horvath, A. L. *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds*; Elsevier: Amsterdam, 1992.
- Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, 69, 17–20.
- Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996**, 100, 10400–10407.
- Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure-Property Relationship. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 28–41.
- Plavsic, D.; Trinajstić, N.; Amic, D.; Soskic, M. Comparison Between the Structure-Boiling Point Relationship with Different Descriptors for Condensed Benzoids. *New J. Chem.* **1998**, 1075–1078.
- Bünz, A. P.; Braun, B.; Janowsky, R. Application of Quantitative Structure-Performance Relationship and Neural Network Models for the Prediction of Physical Properties from Molecular Structure. *Ind. Eng. Chem. Res.* **1998**, 37, 3043–3051.
- Liu, S.; Cao, C.; Li, Z. Approach to Estimation and Prediction for Normal Boiling Point (NBP) of Alkenes Based on a Novel Molecular Distance-Edge (MDE) Vector, λ . *J. Chem. Inf. Comput. Sci.* **1998**, 38, 387–394.
- Grigoros, S. A Structural Approach to Calculate Physical Properties of Pure Organic Substances: The Critical Temperature, Critical Volume and Related Properties. *J. Comput. Chem.* **1990**, 11, 493–510.
- Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. *Anal. Chem.* **1990**, 62, 2323–2329.
- Stanton, D. T.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Furans, Tetrahydrofurans, and Thiophenes. *J. Chem. Inf. Comput. Sci.* **1991**, 31, 301.
- Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Pyrans and Pyrroles. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 306–316.
- Egolf, L. M.; Wessel, M. D.; Jurs, P. C. Prediction of Boiling Points and Critical Temperatures of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 947–956.

- (49) Wessel, M. D.; Jurs, P. C. Prediction of Normal Boiling Points for a Diverse Set of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 841–850.
- (50) Basak, A. C.; Gute, B. D.; Grunwald, G. D. Comparative Study of Topological and Geometrical Parameters in Estimating Normal Boiling Points and Octanol/Water Partition Coefficient. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1054–1060.
- (51) Murugan, R.; Grendze, M. P.; Toomey, J. E.; Katritzky, A. R.; Karelson, M.; Lobanov, V.; Rachwal, P. Predicting Physical Properties from Molecular Structure. *CHEMTECH* **1994**, *24*, 17–23.
- (52) Katritzky, A. R.; Lobanov, V. S.; Karelson, M.; Murugan, R.; Grendze, M. P.; Toomey, J. E., Jr. Comprehensive Descriptors for Structural and Statistical Analysis. 1. Correlations Between Structure and Physical Properties of Substituted Pyridines. *Rev. Roum. Chim.* **1996**, *41*, 851–867.
- (53) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative Structure-Property Relationship Study of Normal Boiling points for Halogen-/Oxygen-/Sulfur-Containing Organic Compounds Using the CODESSA Program. *Tetrahedron* **1998**, *54*, 9129–9142.
- (54) Cherqaoui, D.; Villemain, D. Use of a Neural Network to Determine the Boiling Point of Alkanes. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 97–102.
- (55) Cherqaoui, D.; Villemain, D.; Mesbah, A.; Cense, J.-M.; Kvasnicka, V. Use of a Neural Network to Determine the Boiling Point of Acyclic Ethers, Peroxides, Acetals and their Sulfur Analogues. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 2015–2019.
- (56) Balaban, A. T.; Kier, L. B.; Joshi, N. Correlations between Chemical Structure Normal Boiling Points of Acyclic Ethers, Peroxides, Acetals, and Their Sulfur Analogues. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 237–247.
- (57) Ego, L. M.; Jurs, P. C. Prediction of Boiling Points of Organic Heterocyclic Compounds Using Regression and Neural Network Techniques. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 616–625.
- (58) Zhang, R.; Liu, S.; Liu, M.; Hu, Z. Neural Network-Molecular Descriptors Approach to the Prediction of Properties of Alkenes. *Comput. Chem.* **1997**, *21*, 335–341.
- (59) Liu, S.; Zhang, R.; Liu, M.; Hu, Z. Neural Network-Topological Indices Approach to the Prediction of Properties of Alkenes. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1146–1151.
- (60) Hall, L. H.; Story, C. T. Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electrotological State Indices Using Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004–1014.
- (61) Wessel, M. D.; Jurs, P. C. Prediction of Normal Boiling Points of Hydrocarbons from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 68–76.
- (62) Gakh, A. A.; Gakh, E. G.; Stumper, B. G.; Noid, D. W. Neural Network-Graph Theory Approach to the Prediction of the Physical Properties of Organic Compound. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 832–839.
- (63) Katritzky, A. R.; Mu, L.; Karelson, M. Relationship of Critical Temperatures to Calculated Molecular Properties. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 293–299.
- (64) Turner, B. E.; Costello, C. L.; Jurs, P. C. Prediction of Critical temperatures and Pressure of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 639–645.
- (65) Politzer, P.; Truhlar, D. G. *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Plenum Press: New York, 1981.
- (66) Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, E.; Politzer, P. Relationship of Critical Constants and Boiling Points to Computed Molecular Surface Properties. *J. Phys. Chem.* **1993**, *97*, 9369–9373.
- (67) Katritzky, A. R.; Wang, Y.; Sild, S.; Tamm, T.; Karelson, M. QSPR Studies on Vapor Pressure, Aqueous Solubility and the Prediction of Water-Air Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 720–725.
- (68) Chastrette, M.; Cretin, D.; Tiyal, F.; Structure-Property Relationships: Determination of the Vapour Pressure of Alkanes and Alcohols Using Multifunctional Autocorrelation Method (MAM). *C. R. Acad. Sci. II* **1994**, *318*, 1059–1065.
- (69) Basak, S. C.; Gute, B. D.; Grunwald, G. D. Use of Topostructural, and Topochemical, and Geometrical Parameters in the Prediction of Vapor Pressure: A Hierarchical QSAR Approach. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 651–655.
- (70) Liang, C.; Gallagher, D. A. QSPR Prediction of Vapor Pressure from Solely Theoretically-Derived Descriptors. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 321–324.
- (71) Ego, L. M.; Jurs, P. C. Estimation of Autignition Temperatures of Hydrocarbons, Alcohols, and Esters from Molecular Structure. *Ind. Eng. Chem. Res.* **1992**, *31*, 1798–1807.
- (72) Mitchell, B. E.; Jurs, P. C. Prediction of Autoignition Temperatures of Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 538–547.
- (73) Tetteh, J.; Metcalfe, E.; Howells, S. L. Optimization of radial basis and backpropagation neural networks for modelling auto-ignition temperature by quantitative-structure property relationships. *Chemometr. Intell. Lab. Syst.* **1996**, *32*, 177–191.
- (74) Tetteh, J.; Howells, S.; Metcalfe, E.; Suzuki, T. Optimization of Radial Basis Function Neural Networks Using Biharmonic Spline Interpolation. *Chemometr. Intell. Lab.* **1998**, *41*, 17–29.
- (75) Karelson, M.; Perkson, A. The Correlation and Prediction of Densities of Organic Liquids, *Comput. Chem.* **1999**, *23*, 49–59.
- (76) Katritzky, A. R.; Sild, S.; Karelson, M. A General QSPR Treatment of the Refractive Index of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 840–844.
- (77) Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 913–919.
- (78) Charton, M.; Charton, B.; Quantitative Description of Structural Effects on Melting Points of Substituted Alkanes. *J. Phys. Org. Chem.* **1994**, *7*, 196–206.
- (79) Gramatica, P.; Navas, N.; Todeschini, R. 3D-Modelling and Prediction by WHIM Descriptors. Part 9. Chromatographic Relative Retention Time and Physico-Chemical Properties of Polychlorinated Biphenyls (PCBs). *Chemometr. Intell. Lab.* **1998**, *40*, 53–63.
- (80) Leo, A. J. Calculating log P_{oct} from Structures. *Chem. Rev.* **1993**, *93*, 1281–1306.
- (81) Sangster, J. *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*; John Wiley: Chichester, 1997; Vol. 2.
- (82) Mannhold, R.; Rekker, R. F.; Dross, K.; Bijloo, G.; de Vries, G. The Lipophilic Behaviour of Organic Compounds: 1. An Updating of the Hydrophobic Fragmental Constant Approach. *Quant. Struct.-Act. Relat.* **1998**, *17*, 517–536.
- (83) Buchwald, P.; Bodor, N. Octanol-Water Partition: Searching for Predictive Models. *Curr. Med. Chem.* **1998**, *5*, 353–380.
- (84) Fujita, T.; Iwasa, J.; Hansch, C. A New Substituent Constant, π , Derived From Partition Coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180.
- (85) Nys, G. G.; Rekker, R. F. Statistical Analysis of a Series of Partition Coefficients with Special Reference to the Predictability of Folding of drug Molecules. Introduction of Hydrophobic Fragmental Constants (f-values). *Chim. Ther.* **1973**, *8*, 521–535.
- (86) Nys, G. G.; Rekker, R. F. The Concept of Hydrophobic Fragmental Constants (f-values). II. Extension of its Applicability to the Calculation of Lipophilicities of Aromatic and Hetero-Aromatic Structures. *Chim. Ther.* **1974**, *9*, 361–375.
- (87) Rekker, R. F.; de Kort, H. M. The Hydrophobic Fragmental Constant; An Extension to a 1000 Data Point Set. *Eur. J. Med. Chem.* **1979**, *14*, 479–488.
- (88) Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity. The Hydrophobic Fragmental Constant Approach*; VCH: Weinheim 1992.
- (89) Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley: New York, 1979.
- (90) Leo, A. J. In *Comprehensive Medicinal Chemistry*; Ramsden, C. A., Ed.; Pergamon Press: Oxford, 1990; Vol. 4, pp 295–319.
- (91) Broto, P.; Moreau, G.; Vandycke, C. Molecular Structures, Perception, Autocorrelation Descriptor and SAR Studies; System of Atomic Contributions for the Calculation of the Octanol-Water Partition Coefficient. *Eur. J. Med. Chem.* **1984**, *19*, 71–78.
- (92) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity relationships. I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565–577.
- (93) Ghose, A. K.; Pritchett, A.; Crippen, G. M. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity relationships. III. Modeling Hydrophobic Interactions. *J. Comput. Chem.* **1988**, *9*, 80–90.
- (94) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- (95) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Prediction of Hydrophobic (Lypophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102*, 3762–3772.
- (96) Suzuki, T.; Kudo, Y. Automated log P Estimation Based on Combined Additive Modeling Methods. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 155–198.
- (97) Wang, R.; Fu, Y.; Lai, L. A New Atom-Additive Method for Calculating Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615–621.

- (98) Klopman, G.; Wang, S. A Computer Automated Structure Evaluation (CASE) Approach to Calculation of Partition Coefficient. *J. Comput. Chem.* **1991**, *12*, 1025–1032.
- (99) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. Computer Automated log P Calculation Based on Extended Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752–781.
- (100) Meylan, W. M.; Howard, P. H. Atom/Fragment Contribution Method for Estimating Octanol-Water Partition Coefficients. *J. Pharm. Sci.* **1995**, *84*, 83–92.
- (101) Kamlet, M. J.; Abboud, J. L.; Taft, R. W. The Solvatochromic Comparison Method. 6. The π^* Scale of Solvent Polarities. *J. Am. Chem. Soc.* **1977**, *99*, 6027–6038.
- (102) Kamlet, M. J.; Abboud, J. L.; Abraham, M. H.; Taft, R. W. Linear Solvation Energy Relationship. 23. A Comprehensive Collection of the Solvatochromic Parameters, π^* , α and β and Some Methods for Simplifying the Generalized Solvatochromic Equation. *J. Org. Chem.* **1983**, *48*, 2877–2887.
- (103) Kamlet, M. J.; Doherty, R. M.; Carr, P.; Abraham, M. H.; Marcus, Y.; Taft, R. W. Linear Solvation Energy Relationships. 46. An Improved Equation for Correlation and Prediction of Octanol/Water Partition Coefficients of Organic Non-Electrolytes (Including Strong Hydrogen Bond Donor Solutes). *J. Phys. Chem.* **1988**, *92*, 5244–5255.
- (104) Luehrs, D. C.; Hickey, J. P.; Godbole, K. A.; Rogers, T. N. Estimation of Octanol/Water Partition Coefficients Using LSER Parameters. *Adv. Environ. Res.* **1998**, *2*, 181–185.
- (105) Famini, G. R. *Using Theoretical Descriptors in Quantitative Structure-Activity Relationships*; V.CRDEC-TR-085; U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, 1989.
- (106) Famini, G. R.; Penski, C. A.; Wilson, L. Y. Using Theoretical Descriptors in Quantitative Structure Activity Relationships: Some Physicochemical Properties. *J. Phys. Org. Chem.* **1992**, *5*, 395–405.
- (107) Camilleri, P.; Watts, S. A.; Boraston, J. A. A Surface Area Approach to Determination of Partition Coefficients. *J. Chem. Soc., Perkin Trans. 2* **1988**, 1699–1707.
- (108) Masuda, T.; Jikihara, T.; Nakamura, K.; Kimura, A.; Takagi, T.; Fujiwara, H. Introduction of Solvent-Accessible Surface Area in the Calculation of the Hydrophobicity Parameter log P from an Atomistic Approach. *J. Pharm. Sci.* **1997**, *86*, 57–63.
- (109) Brinck, T.; Murray, J. S.; Politzer, P. Octanol/Water Partition Coefficient Expressed in Terms of Solute Molecular Surface Areas and Electrostatic Potentials. *J. Org. Chem.* **1993**, *58*, 7070–7073.
- (110) Murray, J. S.; Brinck, T.; Lane, P.; Paulsen, K.; Politzer, P. Statistically-Based Interaction Indexes Derived from Molecular-Surface Electrostatic Potentials - A General Interaction Properties Function (GIPF). *J. Mol. Struct. THEOCHEM* **1994**, *307*, 55–64.
- (111) Du, Q.; Arteca, G. A. Modeling Lipophilicity from the Distribution of Electrostatic Potential on a Molecular Surface. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 133–144.
- (112) Sasaki, Y.; Kubodera, H.; Matuszaki, T.; Umeiyama, H. Prediction of Octanol/Water Partition Coefficients Using Parameters Derived from Molecular Structures. *J. Pharmacobiol.-Dyn.* **1991**, *14*, 207–214.
- (113) Klopman, G.; Iroff, L. D. Calculation of Partition Coefficients by the Charge Density Method. *J. Comput. Chem.* **1981**, *2*, 157–160.
- (114) Klopman, G.; Namboodiri, K.; Schochet, M. Simple Method of Computing the Partition Coefficient. *J. Comput. Chem.* **1985**, *6*, 28–38.
- (115) Bodor, N.; Gabanyi, Z.; Wong, C.-K. A New Method for the Estimation of Partition Coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786.
- (116) Bodor, N.; Huang, M. J. An Extended Version of a Novel Method for the Estimation of Partition Coefficients. *J. Pharm. Sci.* **1992**, *81*, 272–281.
- (117) Kantola, A.; Villar, H. O.; Loew, G. H. Atom Based Parametrization for a Conformationally Dependent Hydrophobic Index. *J. Comput. Chem.* **1991**, *12*, 681–689.
- (118) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127–130.
- (119) Moriguchi, I.; Hirono, S.; Nakagome, I.; Hirano, H. Comparison of Reliability of log P Values for Drugs Calculated by Several Methods. *Chem. Pharm. Bull.* **1994**, *42*, 976–978.
- (120) Grombar, V. K.; Enslein, K. Assessment of n-Octanol/Water Partition Coefficient: When Is the Assessment Reliable? *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1127–1134.
- (121) Bodor, N.; Buchwald, P. Molecular Size Based Approach To Estimate Partition Properties for Organic Solutes. *J. Phys. Chem. B* **1997**, *101*, 3404–3412.
- (122) Edward, J. T. Calculation of Octanol-Water Partition Coefficients of Organic Solutes from Their Molecular Volumes. *Can. J. Chem.* **1998**, *76*, 1294–1303.
- (123) Waller, C. L. A Three-Dimensional Technique for the Calculation of Octanol-Water Partition Coefficients. *Quant. Struct.-Act. Relat.* **1994**, *13*, 172–176.
- (124) Grunenberg, J.; Herges, R. Prediction of Chromatographic Retention Values (RM) and Partition Coefficients (log P_{oct}) Using a Combination of Semiempirical Self-Consistent Reaction Field Calculations and Neural Networks. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 905–911.
- (125) Breindl, A.; Beck, B.; Clark, T.; Glen, R. C. Prediction of the n-Octanol/Water Partition Coefficient, log P, Using a Combination of Semiempirical MO-calculations and a Neural Network. *J. Mol. Model.* **1997**, *3*, 142–155.
- (126) Devillers, J. *Neural Networks in QSAR and Drug Design*; Academic Press: London, 1996.
- (127) Rekker, R. F.; ter Laak, A. N.; Mannhold, R. On the Reliability of Calculated log P-values: Rekker, Hansch/Leo and Suzuki Approach. *Quant. Struct.-Act. Relat.* **1993**, *12*, 152–157.
- (128) Mannhold, R.; Rekker, R. F.; Sonntag, C.; ter Laak, A. N.; Dross, K.; Polymeropoulos, E. Comparative Evaluation of the Predictive Power of Calculation Procedures for Molecular Lipophilicity. *J. Pharm. Sci.* **1995**, *84*, 1410–1419.
- (129) Mannhold, R.; Dross, K. Calculation Procedures for Molecular Lipophilicity: a Comparative Study. *Quant. Struct.-Act. Relat.* **1996**, *15*, 403–409.
- (130) Hansch, C.; Quinn, J. E.; Lawrence, G. L. The Linear Free-Energy Relationship between Partition Coefficients and the Aqueous Solubility of Organic Liquids. *J. Org. Chem.* **1968**, *33*, 347–350.
- (131) Banerjee, S.; Yalkowsky, S. H.; Valvani, S. C. Water Solubility and Octanol/Water Partition Coefficients of Organics. Limitations of the Solubility-Partition Coefficient Correlation. *Environ. Sci. Technol.* **1980**, *14*, 1227–1229.
- (132) Mackay, D.; Bobra, A.; Shiu, W. Y.; Yalkowsky, S. H. Relationships between Aqueous Solubility and Octanol-Water Partition Coefficients. *Chemosphere* **1980**, *9*, 701–712.
- (133) Miller, M. M.; Wasik, S. P.; Huang, G. L.; Shiu, W. Y.; Mackay, D. Relationships between Octanol-Water Partition Coefficient and Aqueous Solubility. *Environ. Sci. Technol.* **1985**, *19*, 522–529.
- (134) Yalkowsky, S. H.; Valvani, S. C. Solubilities and Partitioning 2. Relationships between Aqueous Solubilities, Partition Coefficients, and Molecular Surface Areas of Rigid Aromatic Hydrocarbons. *J. Chem. Eng. Data* **1979**, *24*, 127–129.
- (135) Isnard, P.; Lambert, S. Aqueous Solubility and n-Octanol/Water Partition Coefficient Correlations. *Chemosphere* **1989**, *18*, 1837–1853.
- (136) Krop, H. B.; van Velzen, M. J. M.; Parsons, J. R.; Govers, H. A. J. n-Octanol-Water Partition Coefficients, Aqueous Solubilities, and Henry's Law Constants of Fatty Acid Esters. *Chemosphere* **1997**, *34*, 107–119.
- (137) Yalkowsky, S. H.; Banerjee, S. *Aqueous Solubility. Methods of Estimation for Organic Compounds*; Dekker: New York, 1992.
- (138) Yalkowsky, S. H.; Orr, R. J.; Valvani, S. C. Solubility and Partitioning. 3. The Solubility of Halobenzenes in Water. *Ind. Eng. Chem. Fundam.* **1979**, *18*, 351–353.
- (139) Dunnivant, F. M.; Elzerman, A. W.; Jurs, P. C.; Hasan, M. N. Quantitative Structure-Property Relationships for Aqueous Solubilities and Henry's Law Constants of Polychlorinated Biphenyls. *Environ. Sci. Technol.* **1992**, *26*, 1567–1573.
- (140) Amidon, G. L.; Yalkowsky, S. H.; Anik, S. T.; Valvani, S. C. Solubility of Nonelectrolytes in Polar Solvents. V: Estimation of the Solubility of Aliphatic Monofunctional Compounds in Water Using a Molecular Surface Area Approach. *J. Phys. Chem.* **1975**, *79*, 2239–2246.
- (141) Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Carr, P. W.; Doherty, R. F.; Taft, R. W. Linear Solvation Energy Relationships. 41. Important Differences between Aqueous Solubility Relationships for Aliphatic and Aromatic Solutes. *J. Phys. Chem.* **1987**, *91*, 1996–2004.
- (142) Yalkowsky, S. H. Estimation of the aqueous solubility of complex organic compounds. *Chemosphere* **1993**, *26*, 1239–1261.
- (143) Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of aqueous solubility of organic molecules by the group contribution approach. Application to the study of biodegradation. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 474–482.
- (144) Suzuki, T. Development of an Automatic Estimation System for Both the Partition Coefficient and Aqueous Solubility. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 149–166.
- (145) Myrdal, P. B.; Manka, A. M.; Yalkowsky, S. H. AQUAFAC 3: Aqueous Functional Group Activity Coefficients; Application to the Estimation of Aqueous Solubility. *Chemosphere* **1995**, *30*, 1619–1637.
- (146) Nirmalakhandan, N. N.; Speece, R. E. Prediction of Aqueous Solubility of Organic Chemicals Based on Molecular Structure. *Environ. Sci. Technol.* **1988**, *22*, 328–338.

- (147) Li, A.; Doucette, W. J.; Andren, A. W. Estimation of Aqueous Solubility, Octanol/Water Partition Coefficient, and Henry's Law Constant for Polychlorinated Biphenyls Using UNIFAC. *Chemosphere* **1994**, 29, 657–669.
- (148) Kan, A. T.; Tomson, M. B. UNIFAC Prediction of Aqueous and Nonaqueous Solubilities of Chemicals with Environmental Interest. *Environ. Sci. Technol.* **1996**, 30, 1369–1376.
- (149) Patil, G. S. Correlation of aqueous solubility and octanol-water partition coefficient based on molecular structure. *Chemosphere* **1991**, 22, 723–738.
- (150) Patil, G. S. Prediction of aqueous solubility and octanol-water partition coefficient for pesticides based on their molecular structure. *J. Hazard. Mater.* **1994**, 36, 35–43.
- (151) Bodor, N.; Huang, M.-J. A New Method for the Estimation of the Aqueous Solubility of Organic Compounds. *J. Pharm. Sci.* **1992**, 81, 954–960.
- (152) Nelson, T. M.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 601–609.
- (153) Sutter, J. M.; Jurs, P. C. Prediction of Aqueous Solubility for a Diverse Set of Heteroatom-Containing Organic Compounds Using a Quantitative Structure-Property Relationship. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 100–107.
- (154) Huibers, P. H. T.; Katritzky, A. R. Correlation of Aqueous Solubility of Hydrocarbons and Halogenated Hydrocarbons with Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 283–292.
- (155) Ruelle, P.; Kesselring, U. W. The Hydrophobic Propensity of Water toward Amphiprotic Solutes: Prediction and Molecular Origin of the Aqueous Solubility of Aliphatic Alcohols. *J. Pharm. Sci.* **1997**, 86, 179–186.
- (156) Ruelle, P.; Kesselring, U. W. Prediction of the Aqueous Solubility of Proton-Acceptor Oxygen-Containing Compounds by the Mobile Order Solubility Model. *J. Chem. Soc., Faraday Trans.* **1997**, 93, 2049–2052.
- (157) Nouwen, J.; Hansen, B. Correlation Analysis Between Watersolubility, Octanol-water Partition Coefficient and Melting Point Based on Clustering. *Quant. Struct.-Act. Relat.* **1996**, 15, 17–30.
- (158) Bodor, N.; Harget, A.; Huang, M. Neural Network Studies. 1. Estimation of the Aqueous Solubility of Organic Compounds. *J. Am. Chem. Soc.* **1991**, 113, 9480–9483.
- (159) Bodor, N.; Huang, M.; Harget, A. Neural Network Studies. 4. An Extended Study of the Aqueous Solubility of Organic Compounds. *International Journal of Quantum Chemistry: Quantum Chemistry Symposium Vol. 26*; John Wiley & Sons: New York, 1992; pp 853–867.
- (160) Huuskonen, J.; Salo, J.; Taskinen, J. Neural Network Modeling for Estimation of the Aqueous Solubility of Structurally Related Drugs. *J. Pharm. Sci.* **1997**, 86, 450–454.
- (161) Huuskonen, J.; Salo, J.; Taskinen, J. Aqueous Solubility Prediction of Drugs Based on Molecular Topology and Neural Network Modeling. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 450–456.
- (162) Mitchell, B. E.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 489–496.
- (163) Mackay, D.; Shiu, W. Y. A Critical Review of Henry's Law Constants for Chemicals of Environmental Interest. *J. Phys. Chem. Ref. Data* **1981**, 10, 1175–1199.
- (164) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors that Influence the Solubility of Gases and Vapours in Water at 298 K, and a New Method for its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.
- (165) Staudinger, J.; Roberts, P. V. A Critical Review of Henry's Law Constants for Environmental Applications. *Crit. Rev. Environ. Sci. Technol.* **1996**, 26, 205–297.
- (166) Hine, H.; Mookerjee, P. K. The Intrinsic Hydrophilic Character of Organic Compounds. Correlations in Terms of Structural Contributions. *J. Org. Chem.* **1975**, 40, 292–298.
- (167) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. Group Contributions to the Thermodynamic Properties of Non-Ionic Organic Solutes in Dilute Aqueous Solution. *J. Solution Chem.* **1981**, 10, 563–595.
- (168) Nirmalakhandan, N. N.; Speece, R. E. QSAR Model for Predicting Henry's Constant. *Environ. Sci. Technol.* **1988**, 22, 1349–1357.
- (169) Horvath, A. L. *Halogenated Hydrocarbons*; Marcel Dekker: New York, 1982.
- (170) Russell, C. J.; Dixon, S. L.; Jurs, P. C. Computer-Assisted Study of the Relationship Between Molecular Structure and Henry's Law Constant. *Anal. Chem.* **1992**, 64, 1350–1355.
- (171) Katritzky, A. R.; Mu, L.; Karelson, M. A QSPR Study of the Solubility of Gases and Vapors in Water. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 1162–1168.
- (172) Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed.; VCH: Weinheim, 1988.
- (173) Koppel, I. A.; Palm, V. A. The Influence of the Solvent on Organic Reactivity. In *Advances in Linear Free Energy Relationships*; Chapman, N. B., Shorter, J., Eds.; Plenum Press: London, 1972; pp 203–280.
- (174) Sandström, M.; Persson, I.; Persson, P. A Study of Solvent Electron-Pair Donor Ability and Lewis Basicity Scales. *Acta Chem. Scand.* **1990**, 44, 653–675.
- (175) Marcus, Y. The Properties of Organic Liquids that are Relevant to their Use as Solvating Solvents. *Chem. Soc. Rev.* **1993**, 409–416.
- (176) Politzer, P.; Murray, J. S., Eds. Quantitative Treatments of Solute/Solvent Interactions *Theoretical and Computational Chemistry*; Elsevier: New York, 1994.
- (177) Katritzky, A. R.; Mu, L.; Karelson, M. QSPR Treatment of the Unified Nonspecific Solvent Polarity Scale. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 756–761.
- (178) Mu, L.; Drago, R. S.; Richardson, D. E. A Model Based QSPR Analysis of the Unified non-specific solvent polarity scale. *J. Chem. Soc., Perkin Trans. 2* **1998**, 159–167.
- (179) Katritzky, A. R.; Tamm, T.; Wang, Y.; Sild, S.; Karelson, M. QSPR Treatment of Solvent Polarity Scale. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 684–691.
- (180) Katritzky, A. R.; Tamm, T.; Wang, Y.; Karelson, M. A Unified Treatment of Solvent Polarity. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 692–698.
- (181) Eiceman, A. G.; Herbert, H. H., Jr. Gas Chromatography. *Anal. Chem.* **1998**, 70, 321R–339R.
- (182) Bartle, K. D.; Lee, M. L.; Wise, S. A. Factors Affecting the Retention of Polycyclic Aromatic Hydrocarbons in Gas Chromatography. *Chromatographia* **1981**, 14, 69–72.
- (183) Whalen-Pedersen, E. K.; Jurs, P. C. Calculation of Linear Temperature Programmed Capillary Gas Chromatographic Retention Indices of Polycyclic Aromatic Compounds. *Anal. Chem.* **1981**, 53, 2184–2187.
- (184) Doherty, P. J.; Hoes, R. M.; Robbat, A., Jr.; White, C. M. Relationship between Gas Chromatographic Retention Indices and Molecular Connectivities of Nitrated Polycyclic Aromatic Hydrocarbons. *Anal. Chem.* **1984**, 56, 2697–2701.
- (185) Mihara, S.; Enomoto, H. Calculation of Retention Indices of pyrazines on the basis of molecular structure. *J. Chromatogr.* **1985**, 324, 428–430.
- (186) Osmialowski, K.; Halkiewicz, J.; Radecki, A.; Kaliszan, R. Quantum Chemical Parameters in Correlation Analysis of Gas-Liquid Chromatographic Retention Indices of Amines. *J. Chromatogr.* **1985**, 346, 53–60.
- (187) Osmialowski, K.; Halkiewicz, J.; Kaliszan, R. Quantum Chemical Parameters in Correlation Analysis of Gas-Liquid Chromatographic Retention Indices of Amines. II. Topological electronic index. *J. Chromatogr.* **1986**, 361, 63–69.
- (188) Rohrbach, R. H.; Jurs, P. C. Prediction of Gas Chromatographic Retention Indexes of Polycyclic Aromatic Compounds and Nitrated Polycyclic Aromatic Compounds. *Anal. Chem.* **1986**, 58, 1210–1212.
- (189) Mihara, S.; Masuda, H. Correlation Between Molecular Structures and Retention Indices of Pyrazines. *J. Chromatogr.* **1987**, 402, 309–317.
- (190) Rohrbach, R. H.; Jurs, P. C. Prediction of Gas Chromatographic Retention Indexes for Diverse Drug Compounds. *Anal. Chem.* **1988**, 60, 2249–2253.
- (191) Robbat, A.; Jr., Xyrafas, G.; Marshall, D. Prediction of Gas Chromatographic Retention Characteristic of Polychlorinated Biphenyls. *Anal. Chem.* **1988**, 60, 982–985.
- (192) Heinzen, V. E. F.; Yunes, R. A. Using Topological Indices in the Prediction of Gas Chromatographic Retention Indices of Linear Alkylbenzene Isomers. *J. Chromatogr., A* **1996**, 719, 462–467.
- (193) Payares, P.; Diaz, D.; Olivero, J.; Vivas, R.; Gomez, I. Prediction of the Gas Chromatographic Retention Times of Flavonoids from Molecular Structure. *J. Chromatogr., A* **1997**, 771, 213–219.
- (194) Zenkevich, I. G. Principle of Structural Analogy in the Calculation of Gas-Chromatographic Retention Indexes from Physico-Chemical Properties of Organic Compounds. *J. Anal. Chem.* **1998**, 53, 35–40.
- (195) Michotte, Y.; Massart, D. L. Molecular connectivity and retention indexes. *J. Pharm. Sci.* **1977**, 66, 1630–1632.
- (196) Bonchev, D.; Mekenjan, O.; Protic, G.; Trinajstić, N. Application of Topological Indices to Gas Chromatographic Data: Calculation of the Retention Indices of Isomeric Alkylbenzenes. *J. Chromatogr.* **1979**, 176, 149–156.
- (197) Kier, L. B.; Hall, L. H. Molecular Connectivity Analysis of Structure Influencing Chromatographic Retention Indices. *J. Pharm. Sci.* **1979**, 68, 120–122.
- (198) Duvenbeck, Ch.; Zinn, P. List Operations on Chemical Graphs. 3. Development of Vertex and Edge Models for Fitting Retention Index Data. *J. Chem. Inf. Comput. Sci.* **1993**, 33, 211–219.

- (199) Duvenbeck, Ch.; Zinn, P. List Operations on Chemical Graphs. 4. Using Edge Models for Prediction of Retention Index Data. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 220–230.
- (200) Stanton, D. T.; Jurs, P. C. Computer-Assisted Prediction of Gas Chromatographic Retention Indices of Pyrazines. *Anal. Chem.* **1989**, *61*, 1328–1332.
- (201) Georgakopoulos, C. G.; Kiburis, J. C.; Jurs, P. C. Prediction of Gas Chromatographic Relative Retention Times of Stimulants and Narcotics. *Anal. Chem.* **1991**, *63*, 2021–2024.
- (202) Georgakopoulos, C. G.; Tsika, O. G.; Kiburis, J. C.; Jurs, P. C. Prediction of Gas-Chromatographic Relative Retention Times of Anabolic-Steroids. *Anal. Chem.* **1991**, *63*, 2025–2028.
- (203) Buydens, L.; Massart, D. L.; Geerlings, P. Prediction of Gas Chromatographic Retention Indexes with Topological, Physicochemical, and Quantum-Chemical Parameters. *Anal. Chem.* **1983**, *55*, 738–744.
- (204) Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Quantitative Structure-Property Relationship Treatment. *Anal. Chem.* **1994**, *66*, 1799–1807.
- (205) Lucic, B.; Trinajstić, N.; Sild, S.; Karelson, M.; Katritzky, A. R. A New Efficient approach for Variable Selection Based on Multi-regression: Prediction of Gas Chromatographic Retention Times and Response Factors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610–621.
- (206) Donovan W. H.; Famini, G. R. Using Theoretical Descriptors in Structure-Activity Relationships: Retention Indices of Sulfur Vesicans and Related Compounds. *J. Chem. Soc., Perkin Trans. 2* **1996**, 83–89.
- (207) Woloszyn, T. F.; Jurs, P. C. Quantitative Structure-Retention Relationship Studies of Sulfur Vesicans. *Anal. Chem.* **1992**, *64*, 3059–3063.
- (208) Bruchmann, A.; Zinn, P.; Haffer, Chr. M. Prediction of Gas Chromatographic Retention Index Data by Neural Networks. *Anal. Chim. Acta* **1993**, *283*, 869–880.
- (209) Sutter, J. M.; Peterson, T. A.; Jurs, P. C. Prediction of Gas Chromatographic Retention Indices of Alkylbenzenes. *Anal. Chim. Acta* **1997**, *342*, 113–122.
- (210) Pompe, M.; Razinger, M.; Novic, M.; Veber, M. Modelling of Gas Chromatographic Retention Indices Using Counterpropagation Neural Networks. *Anal. Chim. Acta* **1997**, *348*, 215–221.
- (211) Kohonen, T. *Self-Organization and Associative Memory*. Springer-Verlag: Berlin, 1988.
- (212) Pompe, M.; Novic, M. Prediction of Gas-Chromatographic Retention Indices Using Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 59–67.
- (213) Scanlon, J. T.; Willis, D. E. Calculation of Flame Ionization Detector Relative Response Factors Using the Effective Carbon Number Concept. *J. Chromatogr. Sci.* **1985**, *23*, 333–340.
- (214) Jelali-Heravi, M.; Fatemi, M. H. Prediction of Flame Ionization Detector Reponse Factors Using an Artificial Neural Network. *J. Chromatogr., A* **1998**, *825*, 161–169.
- (215) Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure-Property Relationship Approach. 1. Nonionic Surfactants. *Langmuir* **1996**, *12*, 1462–1470.
- (216) Huibers, P. D. T.; Lobanov V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure-Property Relationship Approach. 2. Anionic Surfactants. *J. Colloid Interface Sci.* **1997**, *187*, 113–120.
- (217) Huibers, P. D. T.; Shah, D. O.; Katritzky, A. R. Predicting Surfactant Cloud Point from Molecular Structure. *J. Colloid Interface Sci.* **1997**, *193*, 132–136.
- (218) Bicerano, J. *Prediction of Polymer Properties*, 2nd ed.; Marcel Dekker, Inc.: New York, 1996.
- (219) Katritzky, A. R.; Rachwal, P.; Law, K. W.; Karelson, M.; Lobanov, V. S. Prediction of Polymer Glass Transition Temperatures Using a General Quantitative Structure-Property Relationship Treatment. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 879–884.
- (220) Katritzky, A. R.; Sild, S.; Lobanov, V. S.; Karelson, M. Quantitative Structure-Property (QSAR) Correlation of Glass Transition Temperatures of High Molecular Weight Polymers. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 300–304.
- (221) Camelio, P.; Lazzeri, V.; Waegell, B. QSPR in Polymers: A Straightforward New Approach to Calculate the Glass Transition Temperature. *Polym. Repr. Am. Chem. Soc., Div. Polym. Chem.* **1995**, *36*, 661–662.
- (222) Tan, T. T. M.; Rode, B. M. Molecular modelling of polymers 3. Prediction of Glass Transition Temperatures of Poly(acrylic acid), Poly(methacrylic acid) and polyacrylamide derivatives. *Macromol. Theor. Simulat.* **1996**, *5*, 467–475.
- (223) Cypcar, C. C.; Camelio, P.; Lazzeri, V.; Mathias, L. J.; Waegell, B. Prediction of the Glass Transition Temperature of Multicyclic and Bulky Substituted Acrylate and Methacrylate Polymers Using the Energy, Volume, Mass (EVM) QSPR Model. *Macromolecules* **1996**, *29*, 8954–8959.
- (224) Camelio, P.; Cypcar, C. C.; Lazzeri, V.; Waegell, B. A Novel Approach Towards the Prediction of Glass Transition Temperature: Application of the EVM Model, a Designer QSPR Equation for the Prediction of Acrylate and Methacrylate Polymers. *J. Polym. Sci. Polym. Chem.* **1997**, *35*, 2579–2590.
- (225) Rogers, D.; Hopfinger, A. J. Application of Generic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
- (226) Luke, B. T. Evolutionary Programming Applied to the Development of Quantitative Structure-Activity Relationship and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1279–1287.
- (227) Ulmer, C. W.; Smith, D. A.; Sumpter, B. G.; Noid, D. I. Computational Neural Networks and the Rational Design of Polymeric Materials: the Next Generation Polycarbonates. *Comput. Theor. Polym.* **1998**, *8*, 311–321.
- (228) Joyce, S. J.; Osguthorpe, D. J.; Padgett, J. A.; Price, G. J.; Neural-Network prediction of Glass-Transition Temperatures from Monomer Structure. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 2491–2496.
- (229) Katritzky, A. R.; Sild, S.; Karelson, M. Correlation and Prediction of the Refractive Indices of Polymers by QSPR. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1171–1176.
- (230) Ignatz-Hoover, F.; Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Applications of Semiempirical Molecular Orbital Calculations and Quantitative Structure Property Relations in the Study of Accelerated Sulfur Vulcanization. *Rubber Chem. Technol.*, **1999**, in press.

CI9903206