

A General Treatment of Solubility. 1. The QSPR Correlation of Solvation Free Energies of Single Solutes in Series of Solvents

Alan R. Katritzky,* Alexander A. Oliferenko,[#] Polina V. Oliferenko,[#] Ruslan Petrukhin,[‡] and Douglas B. Tatham[§]

Florida Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, Florida 32611-7200

Uko Maran and Andre Lomaka

Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia

William E. Acree, Jr.

Department of Chemistry, University of North Texas, Denton, Texas 76203-5070

Received June 18, 2003

We present an extended QSPR modeling of solubilities of about 500 substances in series of up to 69 diverse solvents. The models are obtained with our new software package, CODESSA PRO, which is furnished with an advanced variable selection procedure and a large pool of theoretically derived molecular descriptors. The squared correlation coefficients and squared standard deviations (variances) range from 0.837 and 0.1 for 2-pyrrolidone to 0.998 and 0.02 for dipropyl ether, respectively. The predictive power of the models was verified by using the “leave-one-out” cross-validation procedure. The QSPR models presented are suitable for the rapid evaluation of solvation free energies of organic compounds.

BACKGROUND TO THE PRESENT SERIES OF PAPERS

Solubility is of the utmost significance in numerous areas of human endeavor and interest. Solubility in water is fundamental to environmental issues such as pollution, erosion, and mass transfer. Solubility in organic solvents forms much of the basis of the chemical industry. Solubility determines shelf life and cross contamination. It is critically linked to bioavailability and thus to the effectiveness of pharmaceuticals, biodegradation, suitability of gaseous anesthetics, blood substitutes, oxygen carriers, etc. Toxicity is critically dependent on solubility.

Very extensive studies have been carried out on the solubilities of various solute–solvent pairs resulting in diverse theories of solute–solvent interactions that form the basis of our knowledge for the understanding of solubility.¹ These theories are based on concepts ranging from quantitative analysis to statistical mechanics and quantum mechanics. Quantitative treatments of solute–solvent interactions in series of compounds have gained wide attraction and have led to various models for explaining solute–solvent behavior.² Most of this work has involved studying a series of solutes dissolved in a single solvent. There are some instances in which the solubilities of a solute in a series of solvents have been examined, as reviewed elsewhere.^{3,4} Many of the previous studies provide valuable contributions to the

understanding of the general phenomena of solute–solvent interactions. In depth comparisons of published data series have revealed that many gaps exist, which render impossible any general comparison of solvent–solute pairs utilizing only experimental data. Therefore we have proposed the combination of *quantitative structure–property/activity relationship* analysis and subsequent *principal component analysis* for the general treatment of solubility.⁵

A common procedure in quantitative structure–property/activity relationships (QSPR/QSAR) analysis is the application of variable selection methods such as stepwise forward selection,^{6,7} genetic algorithms,^{8,9} and simulated annealing^{10,11} for the reduction of descriptor space in order to keep the only most influential descriptors for the prediction of a property (in the present instance solubility). In this first version of our general treatment of solubility we are still using QSPR (applying forward selection of descriptors scales) essentially as a “black box” method to be able to fill gaps in the experimental data by prediction of the missing values. We are aware that QSPR approaches based on theoretical whole-molecule descriptors and variable selection procedures are sometimes criticized because of difficulties in their physical interpretation.^{12,13} We also acknowledge that a rigorous physical interpretation of the quantitative structure–activity relationships that we have derived is not yet possible.

After predicting missing values to provide a continuous data matrix, we plan to further analyze the matrix with principal component analysis in order to disclose general regularities within the solute–solvent pairs. In this way, it may be possible to derive the intrinsic dimensionality of the general solubility phenomenon and to reveal important

* Corresponding author phone: (352)392-0554; e-mail: katritzky@chem.ufl.edu.

[#] Present address: Department of Chemistry, Moscow, Moscow State University, 119899 Russia.

[‡] Present address: ImClone Systems Inc., 180 Varick Street, New York, NY 10014.

[§] Present address: Alchem Laboratories, Alachua, FL 32615.

constitutional and structural factors responsible for the solvation behavior of chemicals. Such an analysis is of interest and importance both from the theoretical and the applied points of view because most physiological and technological processes occur in solvent media, and solvents exert strong influences on the rates and even directions of these processes.¹⁴

The current paper is the first in a series of publications comprising a general treatment of solubility. It initially develops solubility scales for 69 solvents with 14 or more solutes in each of the 69 scales.

INTRODUCTION

Various research groups have treated two basic types of solubility. Bulk solubility, expressed in terms of concentration (molarity), has been addressed inter alia by Jurs,¹⁵ Yalkowsky,¹⁶ and Acree.¹⁷ More generally, solubility can be considered primarily as a partition characteristic and be treated as the ratio of the liquid- and gas-phase molar concentrations of the solute. Contributions from Abraham,¹⁸ and some of our previous work,¹⁹ used the logarithms of gas-solvent or water-solvent partition coefficients, $\log L$ and $\log P$, respectively. Here, L is the Ostwald solubility coefficient and is usually written in its logarithmic form (eq 1).

$$\log L = \log \left(\frac{C_l}{C_g} \right) \quad (1)$$

The use of partition characteristics seems to be the preferable approach because it explicitly addresses the existence of a vapor-liquid equilibrium and allows solubility to be expressed on the energy scale as the relative free energy of solvation (ΔG_s). At equilibrium conditions and constant temperature, the solvation free energy, ΔG_s (the work taken to transfer one mole of a solute from the gas phase to the solution), is equal to the product of $-RT$ and $\log L$. If one chooses standard states of unit concentration in the gas phase and in solution, ΔG_s is related to the logarithm of the Ostwald solubility coefficient by eq 2.

$$\Delta G_s = -2.3RT \log L = -2.3RT \log \left(\frac{C_l}{C_g} \right) \quad (2)$$

Solvation free energies have been extensively modeled by the groups of Politzer,² Cramer and Truhlar,²⁰ Jorgensen,²¹ and Wendoloski.²² In the present study we use the energy scale of solubility because such a representation of solubility and related phenomena is more appropriate for a wide audience in the fields of QSAR/QSPR, biomolecular simulations, and organic reactivity modeling.

Computational Chemistry Approaches. Modern science proposes rational explanations of the macroscopic solubility phenomenon based on the microscopic properties of matter. Statistical mechanics rigorously links these two realms through the probabilistic treatment of particle ensembles. The application of Kirkwood's continuum solvent approach²³ to nondissociating fluids resulted in a variety of simulation techniques: the most popular are molecular dynamics (MD) and the Monte Carlo (MC) method.²⁴

These techniques can provide theoretical characteristics of solutes, which are useful in liquid-phase equilibria calculations and lead to the prediction of solvation free energies and the effects that a solvent exerts on chemical equilibrium and rates. Recent applications of the such techniques include the work of Duffy and Jorgensen²¹ who predicted the solvation free energies of 68 organic substances in hexadecane (squared correlation coefficient $R^2 = 0.90$, unsigned mean error $|\Delta| = 0.33$ kcal/mol), 85 substances in wet octanol ($R^2 = 0.87$, $|\Delta| = 0.52$ kcal/mol), and 85 substances in water ($R^2 = 0.89$, $|\Delta| = 0.54$ kcal/mol) using MC simulation-derived descriptors.

A conceptually similar but methodologically different approach was developed by Murrey et al.²⁵ who correlated the aqueous solvation free energies of 50 organic compounds with some molecular electrostatic potential (MEP)-derived descriptors ($R^2 = 0.988$, $|\Delta| = 0.27$ kcal/mol).

The explicit inclusion of solvation free energy into the framework of the MO SCF method initiated self-consistent reaction field implementations at the semiempirical^{26,27} and ab initio levels.²⁸⁻³⁰ Practical advances in the prediction of solvation free energies of organic compounds in water and about 90 nonaqueous solvents were reached by Cramer and Truhlar. They included a set of empirical parameters in their SM.x semiempirical solvation models^{20,31} that accounts implicitly for the effects of the first solvation shell. Unsigned mean errors of their calculations for the solvation free energies of 275 neutral solutes in 91 solvents range from 0.43 to 0.46 kcal/mol.

In more recent work³² Torrens provides a semiquantitative approach to estimate solvation free energy. The method is an extension of solvent-dependent conformational analysis and uses the dielectric constant and molecular volume to describe solvent.

The above-mentioned methods, using theoretical tools from the arsenal of computational and quantum chemistry, allow the derivation of rather accurate values of solvation free energies, but these methods are time-consuming and can hardly be applied to the solubility modeling of large biomolecules or to the large-scale modeling of many hundreds of small molecules.

QSAR/QSPR Approaches. Several less sophisticated, but also much less time-consuming, methods based on QSAR/QSPR methodology have been developed recently for the prediction of solvation free energies and for the derivation of partitioning characteristics. This family of methods can be divided into two groups: (i) methods using experimental quantities as descriptors and (ii) those based on purely theoretical descriptors.

Experiment-Based QSAR/QSPR. The *linear solvation energy relationships* (LSER) method is based on multilinear regression (MLR) analysis of the solubilities of solutes in different solvents expressed in the form of different solvent polarity scales or solvatochromic parameters. This method, originally developed by Kamlet and Taft,^{33,34} has been further evolved and applied by Abraham and co-workers¹⁸ and applied to numerous solubility and partitioning problems.

The LSER MLR model includes several characteristics (descriptors) that are used to describe the solvent/solute polarizability, dipolarity, volume, hydrogen bond acidity, and hydrogen bond basicity. The strength of the approach relies on combining all these characteristics into a single model,

thus providing a sound basis to discuss solute–solvent interactions and also the ability to rank each type of interaction for every solute–solvent pair. A limitation is that the characteristics used in the model, which originate from experimental measurements, are often unavailable or incomplete when working with diverse compounds within large databases.

Reliable LSER models were obtained for water³⁵ ($R^2 = 0.998$ and standard deviation, $s = 0.151$) and several nonaqueous solvents. The Ostwald solubility coefficient of 77 gases and vapors in propan-1-ol at 298 K³⁶ were correlated with modified five LSER descriptors giving $R^2 = 0.998$ and standard deviation, $s = 0.12$. This equation suggests that propan-1-ol as a solvent is less dipolar, more acidic, and less basic than methanol or ethanol, but that the differences between the three alcohols are small.

The same model was also used to compare a series of aliphatic alcohols: butan-1-ol, pentan-1-ol, hexan-1-ol, heptan-1-ol, and decan-1-ol.³⁷ For 92 solutes in butan-1-ol the R^2 value was 0.9966 ($s = 0.158$), for 61 solutes in pentan-1-ol the R^2 value was 0.999 ($s = 0.076$), for 46 solutes in hexan-1-ol the R^2 value was 0.9996 ($s = 0.089$), for 38 solutes in heptan-1-ol the R^2 value was 0.9998 ($s = 0.0670$), and for 45 solutes in decan-1-ol the R^2 value was 0.9996 ($s = 0.090$).

Abraham and his collaborators have used LSER methodology to analyze 150 values of gas–chloroform partition coefficients, resulting in a standard deviation of 0.23 log units and a correlation coefficient R^2 of 0.985.³⁸ Based on the equation, they conclude that bulk chloroform (i) is dipolar/polarizable, (ii) has low hydrogen-bond basicity, (iii) is as strong a hydrogen-bond acid as bulk methanol or bulk ethanol, and (iv) exerts its main influence on gaseous solubility by solute–solvent London dispersion interactions. Other solvents studied by Abraham and co-workers include the following: *N*-methylpyrrolidone, *N,N*-dimethylformamide, *N,N*-dimethylacetamide,³⁹ and methylene iodide.⁴⁰

The Mobile Order Theory (MOT) approach, originally developed by Huyskens⁴¹ and widely applied by Acree and co-workers,¹⁷ is based on a thermodynamic treatment of the liquid state. It includes terms to describe the effects that solute–solvent, solvent–solvent, and solute–solute interactions have on the chemical potential of the dissolved solute. MOT assumes that hydrogen-bonded aggregates are formed temporarily without a distinguishable thermodynamic identity. Hydrogen-bonded partner associates are not preserved with time but rather change continuously. The MOT treatment considers equilibria involving the fractions of time during which an amphiphilic proton is in the hydrogen-bonded and nonbonded state. This differs from more conventional thermodynamic approaches that treat equilibrium in terms of discrete chemical species. Depending upon the functional groups present on the solute and solvent molecules, the MOT predictive expression may contain up to six terms and require a priori knowledge of several input stability constants before a prediction can be made. The predictions obtained by MOT approach for solubilities of many solutes in nonelectrolyte solvents will be discussed in more detail in part 2 of the present series.⁴²

Another approach to the prediction of aqueous solubility involving experimentally derived entities was provided by Yalkowsky et al.¹⁶ The so-called General Solubility Equation

of Yalkowsky includes two experimental parameters: the melting point of the solute and its octanol–water partition coefficient ($\log P$). The aqueous solubilities of 150 physiologically active compounds were estimated with an unsigned mean error of 0.43 molarity units ($\log S$).

Theory-Based QSAR/QSPR. These approaches exploit two different paradigms. One involves the derivation of various molecular characteristics (descriptors) solely from molecular structure.⁴³ Depending on the level of consideration, one can use simple constitutional features (numbers of atoms or bonds of a particular class), topological indices,⁴⁴ geometric, electrostatic, quantum chemical, and thermodynamic descriptors.^{45,46} Another popular paradigm of QSPR modeling relies on the concept of the structural additivity of properties. According to this hypothesis, any property in the form of a continuous smooth function can be expanded into a linear function in some predefined structural features such as atoms, bonds, chemically relevant groups, and larger fragments of the molecules. Such fragment-based methods are effectively generalizations of the use of constitutional descriptors.

Quantitative structure–property relationships for the solubilities of gases and vapors in water,¹⁹ in methanol, and in ethanol⁴⁷ have been presented in two of our recent publications. For water, we showed that the solubilities of a diverse set of 406 gases and vapors in water could be correlated with a five-parameter equation with $R^2 = 0.942$. In methanol, the solubilities of 87 gases and vapors gave a four-parameter equation ($R^2 = 0.945$, $R^2_{cv} = 0.938$) that adequately accounted for the solute–solvent interactions in terms of polarizability, dipole moment, hydrogen bonding, and lipophilicity.⁴⁷ The solubilities in ethanol of 61 gases and vapors also yielded a four-parameter equation ($R^2 = 0.969$, $R^2_{cv} = 0.964$), where the solute–solvent intercorrelations, similar to methanol, include electrostatic and hydrogen-bonding interactions.⁴⁷ The descriptors in each of these three models are all derived solely from the chemical structures and do not require any experimental data.

Jurs and co-workers exploit a similar approach to correlate aqueous solubility (in molarity units) with molecular structure using both MLR and artificial neural networks (ANN).⁴⁸ Their MLR model resulted in $R^2 = 0.965$ and $s = 0.638$ mol/dm³, whereas the ANN model provided $s = 0.394$ mol/dm³ for the training set of 295 compounds and R^2 between the experimental and predicted values being 0.987. Applying ANN methodology and electrotopological state descriptors, Tetko and Villa⁴⁹ predicted aqueous molar solubilities of 1291 organic compounds with $R^2 = 0.910$ and $s = 0.62$ mol/dm³. Huuskonen^{50,51} reported that the ANN approach applied to aqueous solubility gave better predictive models as compared to MLR.

Fragment-based or group contribution schemes have been actively pursued by the groups of Klopman^{52,53} and Wendoloski.^{22,54} The most frequently occurring chemical groups are considered and, using information from the theory of functions of a complex variable, Klopman et al. derived contribution coefficients for many groups and successfully correlated them with the aqueous solubilities of 1168 organic compounds ($R^2 = 0.95$, $|\Delta| = 0.5 \log S$). The additive-constitutive approach of Wendoloski et al.²² took the form of two predictive models: (i) the method of molecular holograms (HLOGS) counts molecular substructures and uses them as variables in a partial least-squares regression; (ii)

the second model, named ALOGS, utilizes the atomic constant approach including a thorough classification of atomic types and their contributions. Both models provided good results: $R^2 = 0.941$, $\text{rms} = 0.58$ kcal/mol for HLOGS and $R^2 = 0.960$, $\text{rms} = 0.38$ kcal/mol for ALOGS.

The successful LSER methodology discussed above was combined with quantum chemical calculations and found new realization in the *theoretical linear solvation energy relationship* (TLSER) by Famini et al.^{55,56} In TLSER the experimentally derived solvatochromic parameters were substituted by semiempirical electronic indices such as partial charges on certain atoms, HOMO and LUMO energies, etc.

In the present paper, we present a comprehensive database of solubilities, which was collected from the literature and from original measurements. We also describe newly developed methodology for the fast selection of descriptors in quantitative structure–property (QSPR) analysis. Selected solvent series are analyzed, and new QSPRs are derived which will allow the filling of gaps in tables of experimental solubility values.

DATA

The overall data collected includes 141 solvents and 448 solutes. The number of solutes that the solvents span varies from 226 for hexadecane to less than 3 for several solvents. We now report our results on the modeling of 69 of these solvents, which span from 14 to 226 solutes, with the average number of solutes per solvent being 48. The remaining 72 solvents have less than 14 experimental points available.

A significant portion of the available molar solubility data has been either experimentally measured at 25 °C or collected from various sources (see Table 3 of Supporting Information for references) by one of the authors (W.E.A.). Publications of Abraham^{37,57} are another important source of information, especially with solubilities in hexadecane and alcohols.

The inclusion of the experimental data compiled by Cramer and Truhlar in the framework of their AMSOL project needed special care because these data originated basically from liquid–liquid partition measurements, which implies that the solvents are saturated with water, that is, “wet” solvents.³¹ By contrast, the analytical approaches developed by Abraham and Acree dealt with “dry” solvents. Nevertheless, as was demonstrated by Abraham et al.,⁵⁸ the difference is negligible for those solvents in which little water is dissolved.

Solubility molarity values (or infinite dilution activity coefficients) were recalculated, as necessary, for transformation into the logarithmic Ostwald solubility coefficients ($\log L$) using eq 1 and then into solvation free energies by eq 2. The appropriate vapor pressures and density data were used for this purpose.⁵⁹ A wide range of solvents was used in our correlation analysis: inert, polar aprotic, polar protic, and dipolar aprotic.

QSPR models developed for the 69 solvents are listed in Table 1 along with the corresponding statistical characteristics. To build the representative QSPR models, we selected only those solvents for which at least 14 experimental solubilities are available.

METHODOLOGY

Each descriptor in the comprehensive set implemented in the CODESSA PRO software package⁶⁰ is calculated solely

from molecular structures: directly from the molecular formula in the case of constitutional and topological descriptors and utilizing the molecular 3D geometry for descriptors such as geometrical and CPSA (charged partial surface area). Quantum chemical descriptors are extracted from MOPAC⁶¹ output files and include orbital energies and coefficients (and their combinations), atomic and bond populations, various components of the MOPAC energy partitioning scheme, polarizabilities up to second order, dipole moments, and calculated thermodynamical functions. Computational details are given in the Experimental Section.

$$\Delta G_S = \Delta G_{\text{cavity}} + \Delta G_{\text{el}} + \Delta G_{\text{disp}} + \Delta G_{\text{HB}} \quad (3)$$

Before outlining and discussing our results, it is critically important to establish connections between the property under study and our descriptors. Solvation free energy is believed to be comprised of four main components^{1,62,63} as expressed by the general terms in eq 3, where ΔG_{cavity} is the cavity-formation term, ΔG_{el} is the free energy of electrostatic interactions, ΔG_{disp} depicts dispersion interactions, and ΔG_{HB} is the term arising from the hydrogen bond formation. A more detailed subdivision of the energy contributions is displayed in Figure 1, together with the five main families of CODESSA PRO descriptors. The interplay between these entities is obvious as well as instructive for further discussion of the QSPR results. Some of the solvation free energy components are related to several descriptor families. Thus, the cavity formation term can be satisfactorily modeled with the use of topological and geometrical descriptors, semiempirically derived molecular polarizability, and entropy. Electrostatic and quantum chemical descriptors contribute significantly to the nonspecific solvation through atomic charges, charged surface areas, dipole moments, reactivity indices, etc. Specially designed hydrogen bond descriptors include molecular surface areas that are confined by H-bond donor or acceptor sites as well as those that merely count such sites derived from atomic charge considerations. We conclude that each contribution to the solvation free energy (at the left of Figure 1) has its corresponding counterpart among the descriptors. This ensures that all of the main energy components of a solvation process are covered by the theoretically derived descriptors of CODESSA PRO, thus correlations produced with these descriptors should be substantial.

RESULTS

The correlation results for all the 69 solvents are listed in two tables, Tables 1 and 2. Each entry of Table 1 provides statistical characteristics of the chosen QSPR model, including the chemical name of a solute, the number (N) of experimental points used for the correlation, the number of descriptors (n), the squared correlation coefficient (R^2), the cross-validated squared correlation coefficient (R^2_{cv}), and the variance or squared standard deviation (s^2).

Solvents have been divided into four main classes: (i) inert solvents (hydrocarbons—aliphatic and aromatic in Tables 1 and 2), which are subdivided into 13 alkanes and 6 aromatic hydrocarbons with the corresponding halogenated compounds (3 and 2, respectively) inserted at the end of each section; (ii) protic solvents represented by 17 alcohols plus water and a single diol (ethylene glycol); (iii) polar aprotic solvents

Table 1. Statistical Quality of QSPR Models

no.	solvent	<i>N</i>	<i>n</i>	<i>R</i> ²	<i>R</i> ² _{cv}	<i>s</i> ²	<i>F</i> -crit	no.	solvent	<i>N</i>	<i>n</i>	<i>R</i> ²	<i>R</i> ² _{cv}	<i>s</i> ²	<i>F</i> -crit
Hydrocarbons: Aliphatic															
1	<i>n</i> -pentane	26	4	0.964	0.921	0.046	142	9	methylcyclohexane	16	3	0.943	0.892	0.072	67
2	<i>n</i> -hexane	70	5	0.939	0.922	0.227	196	10	2,2,4-trimethylpentane	45	5	0.972	0.962	0.142	271
3	<i>n</i> -heptane	78	3	0.933	0.926	0.235	343	11	<i>tert</i> -butylcyclohexane	16	3	0.936	0.873	0.081	59
4	<i>n</i> -octane	48	3	0.978	0.969	0.104	640	12	<i>trans</i> -decaline	24	4	0.917	0.866	0.050	53
5	<i>n</i> -nonane	38	5	0.981	0.971	0.106	336	13	squalane	58	5	0.973	0.958	0.080	369
6	<i>n</i> -decane	46	4	0.959	0.944	0.191	240	14	chloroform	107	5	0.907	0.888	0.637	196
7	<i>n</i> -hexadecane	226	5	0.928	0.919	0.186	565	15	carbon tetrachloride	81	4	0.953	0.944	0.183	382
8	cyclohexane	104	6	0.944	0.928	0.269	272	16	1,2-dichloroethane	53	5	0.965	0.949	0.299	257
Hydrocarbons: Aromatic															
17	benzene	85	6	0.926	0.905	0.518	162	21	<i>p</i> -xylene	49	5	0.930	0.891	0.458	115
18	toluene	59	5	0.959	0.947	0.264	250	22	ethyl benzene	54	5	0.969	0.956	0.123	297
19	<i>o</i> -xylene	22	3	0.954	0.933	0.457	123	23	chlorobenzene	46	5	0.973	0.959	0.207	289
20	<i>m</i> -xylene	22	3	0.974	0.964	0.367	227	24	bromobenzene	24	4	0.982	0.973	0.029	253
Protic Solvents – Alcohols															
25	water	177	5	0.931	0.926	0.340	464	34	3-methyl-1-butanol	15	3	0.965	0.892	0.069	61
26	methanol	92	5	0.932	0.912	0.508	239	35	1-pentanol	25	4	0.924	0.877	0.379	61
27	ethanol	74	3	0.933	0.924	0.608	326	36	4-methyl-2-pentanol	15	3	0.957	0.913	0.073	82
28	1-propanol	66	4	0.965	0.959	0.226	416	37	1-hexanol	19	3	0.914	0.861	0.138	53
29	2-propanol	72	4	0.965	0.955	0.344	459	38	2-ethyl-1-hexanol	15	3	0.878	0.790	0.236	26
30	1-butanol	21	3	0.956	0.914	0.157	123	39	1-heptanol	16	3	0.921	0.839	0.170	46
31	2-butanol	19	3	0.933	0.908	0.109	70	40	1-octanol	173	5	0.942	0.937	0.579	548
32	2-methyl-1-propanol	17	2	0.948	0.929	0.090	128	41	benzyl alcohol	28	3	0.982	0.974	0.073	425
33	2-methyl-2-butanol	15	3	0.948	0.927	0.075	66	42	ethylene glycol	14	2	0.920	0.882	0.611	63
Polar Aprotic Solvents – Ethers															
43	dipropyl ether	21	3	0.998	0.997	0.016	2745	46	tetrahydrofuran	55	5	0.982	0.976	0.147	524
44	dibutyl ether	38	4	0.959	0.945	0.299	191	47	1,4-dioxane	54	5	0.965	0.954	0.473	264
45	methyl <i>tert</i> -butyl ether	29	2	0.944	0.932	0.342	219								
Dipolar Aprotic Solvents															
48	methyl acetate	36	4	0.976	0.962	0.220	314	59	acetone	59	5	0.963	0.955	0.437	282
49	ethyl acetate	68	5	0.949	0.925	0.580	230	60	2-butanone	54	5	0.980	0.972	0.131	468
50	butyl acetate	31	4	0.987	0.979	0.060	477	61	cyclohexanone	30	4	0.970	0.961	0.136	205
51	acetonitrile	35	3	0.979	0.973	0.183	478	62	acetophenone	35	4	0.982	0.973	0.189	417
52	1-propanenitrile	31	4	0.984	0.974	0.117	394	63	dimethyl sulfoxide	46	4	0.958	0.937	0.463	229
53	1-butanenitrile	39	5	0.982	0.970	0.104	353	64	aniline	42	5	0.937	0.912	0.426	106
54	adiponitrile	16	3	0.994	0.990	0.004	671	65	pyridine	33	5	0.986	0.975	0.068	391
55	benzonitrile	27	3	0.980	0.974	0.090	385	66	2-pyrrolidone	24	3	0.837	0.788	0.093	34
56	<i>N</i> -methylformamide	33	5	0.974	0.957	0.145	204	67	<i>N</i> -methyl-2-pyrrolidone	16	2	0.994	0.991	0.059	1142
57	<i>N,N</i> -dimethylformamide	52	4	0.990	0.984	0.065	723	68	nitromethane	29	4	0.985	0.974	0.059	401
58	<i>N,N</i> -dimethylacetamide	43	3	0.969	0.959	0.234	410	69	propylene carbonate	28	4	0.917	0.873	0.036	63

are represented by 5 ethers; and (iv) a set of 22 dipolar aprotic solvents including 3 esters, 4 ketones, 5 amides, 5 nitriles, 2 amines, 1 sulfoxide, 1 nitroalkane, and 1 carbonate ester.

Table 2 displays the quantitative–structure/activity relationship equations for all 69 solvents. The equations are written in a linear notation, and a key to the independent variables (descriptors) is given in the Discussion section. The number of descriptors involved in the multiregression analysis varies from a maximum of 6 for benzene to 2 for ethylene glycol. Frequency analysis of the descriptors with respect to the solvents overall is displayed in Figure 2 and with respect to individual solvent groups in Figure 3.

Tables 3 (Supporting Information) collects all numerical data on experimental and predicted solubilities in each of the 69 solvents. Each of these tables contains the unique in-house ID of a solute, its chemical name, experimental and predicted solubilities, and the original literature references to each solubility value.

DISCUSSION

The overall statistical quality of the QSPR models derived is satisfactory. The best correlation is a three-parameter correlation obtained for dipropyl ether (eq 43) on the basis

of 21 experimental data, with $R^2 = 0.998$, $R^2_{cv} = 0.997$, and $s^2 = 0.016$; the least satisfactory is a three-parameter correlation for 2-pyrrolidone, eq 66, characterized by $R^2 = 0.837$, $R^2_{cv} = 0.788$, and $s^2 = 0.093$ ($N = 24$). Just two solvents (2-pyrrolidone and 2-ethyl-1-hexanol) have QSPR models with R^2 less than 0.9. Analysis of the variance, s^2 , shows that the predictive ability of the models ranges from excellent (0.004) in adiponitrile (#54) to admissible (0.608) in the case of ethanol (#27). Only 12 of the 69 models have variances greater than the generally accepted 0.4 kcal/mol value of experimental uncertainty.³¹ We provide comparisons with previous theoretical treatments of solubility for those solvents, for which such data are available. Of the solvents (other than water) studied in the current work, comprehensive QSPR studies have been previously conducted only for 11 solvents: hexadecane,^{64,65} methanol,^{47,66} ethanol,^{47,67} 1-propanol,³⁶ 1-butanol, 1-pentanol, 1-hexanol, 1-heptanol,³⁷ 1-octanol,⁶⁸ *N,N*-dimethylacetamide³⁹ and *N,N*-dimethylformamide.³⁹ For the remaining 57 solvents described below such work is now carried out for the first time.

The QSPR equations of Table 2 show some general regularities in the occurrence of descriptors. The frequency analysis shown in Figure 2 indicates that the contributions from electrostatic, topological, and hydrogen bonding de-

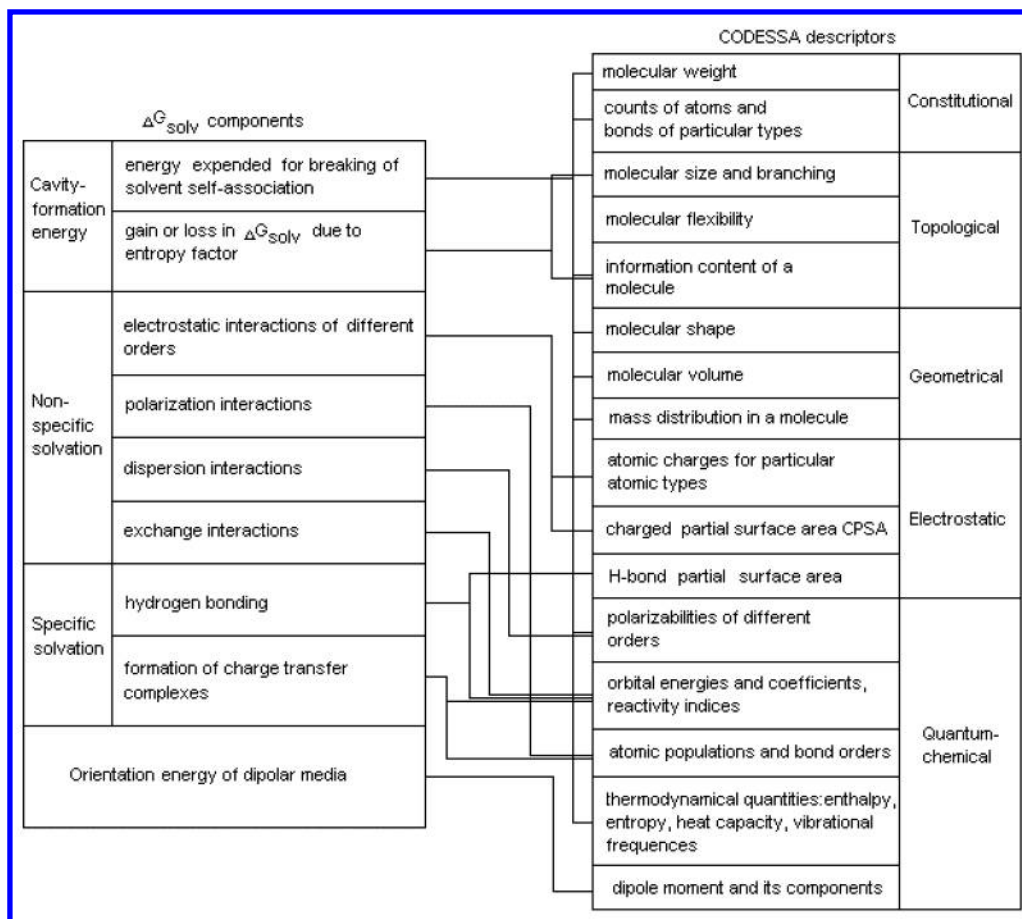


Figure 1. Interplay between solvation free energy components and CODESSA descriptors.

criptors are the most important. Geometrical and quantum chemical descriptors appear to about the same extent, whereas thermodynamic descriptors occur sparsely. A similar trend (Figure 3) is found also within the individual classes of solvents, though within these classes the relative contributions vary. Thus, the solubilities in both aliphatic and aromatic hydrocarbon solvents are basically described by electrostatic and topological descriptors. A possible interpretation is that the highly polarizable hydrocarbon molecules of the solvent media are sensitive to perturbations from electrostatic and dispersion interactions. By contrast, other solubilities are more generally modeled with quantum chemical and hydrogen bonding descriptors, reflecting the polar character and hydrogen bond acceptor nature of ethers. As expected, alcohols are described by topological and hydrogen bonding descriptors, whereas electrostatic descriptors are the most important for modeling solubilities in dipolar aprotic solvents.

We wish to emphasize that our objective in formulating the equations of Table 2 is to allow the estimation of as yet unmeasured solubilities of solutes in solvents. These estimates are needed to complete the matrix for our principal component analysis treatment described in part 3 of this series. Although we have attempted, in the present paper, a preliminary analysis of the descriptors used in the equations in terms of physical interactions, we strongly believe that an improved understanding will be provided when more appropriate descriptors become available. Efforts in this direction are ongoing in our laboratories.

Because of the large amount of the results to be discussed, we have classified them according to solvent classes. A partial key to the descriptors is given below in the text and fully in Table 4 (Supporting Information).

Inert Solvents (i) – Alkanes and Chloroalkanes. As seen in Tables 1 and 2, all the QSPR models for the saturated inert solvents are characterized by good to excellent statistical quality with R^2 varying from 0.907 for chloroform to 0.981 for *n*-nonane. The solvation behavior of alkanes as solvents is revealed from the nature of the relevant descriptors. The low polarity and chemical inertness of alkanes suggest that the main contribution to solvation free energy in alkanes should be cavity-formation and nonspecific solvation terms. In agreement with this, the Randic topological indices⁶⁹ of orders 0 and 1 occur in 7 of the 16 equations.

The Randic index is calculated as a sum of atomic connectivities over molecular paths of certain length (1, 2, ..., *n*), it thus reflects molecular size and branching and can be closely related to the cavity-formation term. Additionally, the Randic index is an implicit measure of dispersion interactions, hence in total it reflects the interplay between two energy terms: (a) the larger the molecular size the stronger the dispersion interactions; (b) other things being equal, the more branched a molecule the bigger its effective radius and thus, the higher the energy cost of the formation of the cavity.

Molecular descriptors such as the gravitational indices calculated over all atoms pairs and bonds, G_p and G_b , respectively, are frequently used in QSPR studies.⁴³ They

Table 2. Final QSPR Models for Solubility of Different Solutes in 69 Solvents

eq	QSPR model
Hydrocarbons: Aliphatic	
1	$\log L = 9.157 - 0.136WNSA^{(2)} + 1.026^0CIC + 0.567\Delta S_{lib}/N_A - 12.365\bar{V}_H$
2	$\log L = -2.455 + 0.0014G_p - 11.729FNSA^{(2)} + 0.778^0CIC + 6.023V_M/V_{XYZ} - 0.026S_{YZ}$
3	$\log L = -1.167 + 0.042\alpha + 0.011\Delta S_{tot} + 1.178N^{occ.el.le}v/N_A$
4	$\log L = 0.768 + 1.126^1\chi + 0.084WNSA^{(3)} - 3.301FPSPA^{(2)}$
5	$\log L = 3.988 + 0.894^0\chi^v + 5.375q_A^{min} - 4.778FPSPA^{(1)} + 0.443^HAHDCA^{(1)} - 0.027PPSA^{(3)}$
6	$\log L = -0.976 + 1.044^1\chi + 0.054S_{ZX} + 0.369^HDCPSA^{(2)} - 0.0032DPSPA^{(1)}$
7	$\log L = -0.700 + 0.0047G_b + 2.325V_M/V_{XYZ} + 0.537T_b^E + 24.295FNSA^{(3)} + 3.737FPSPA^{(3)}$
8	$\log L = 0.072 + 1.083^1\chi - 0.261^2IC + 1.947S_{XY}/R_{XY} + 21.328FNSA^{(3)} - 0.298^1CIC + 0.352N^{HA}$
9	$\log L = 3.651 + 0.606^0\chi - 0.577J + 0.592WNSA^{(3)}$
10	$\log L = -1.164 + 0.929^0\chi + 0.013PNSA^{(3)} + 43.705FPSPA^{(3)} - 0.969FPSPA^{(2)} - 5.156q_C^{max}$
11	$\log L = 3.257 + 0.619^0\chi + 0.577WNSA^{(3)} - 0.482J$
12	$\log L = 6.759 + 0.031S_M - 9.691\bar{V}_H + 24.478^HFCPSA^{(2)} + 0.083\nu_{TD}^h$
13	$\log L = -1.047 + 0.988^1\chi^v + 0.028WNSA^{(1)} + 0.426E_{tot} - 1.628S_{YZ}/R_{YZ} - 2.899q_A^{min}$
14	$\log L = -3.003 + 0.0043G_b + 0.397^2IC + 0.790(HA,HD)^{max/min} + 0.067\Delta H_f^0/N_A + 5.688S_{XY}/R_{XY}$
15	$\log L = 18.174 + 0.049\alpha - 1.252FNSA^{(2)} - 18.996P_H^{min} + 0.142PPSA^{(3)}$
16	$\log L = 0.187 + 1.127^1\chi - 0.0061PPSA^{(2)} - 122.762^HFCPSA^{(2)} - 13.821FNSA^{(3)} + 1.167(HA,HD)^{max/min}$
Hydrocarbons: Aromatic	
17	$\log L = -4.859 + 1.073^1\chi - 0.074^2IC - 0.114\Phi + 8.155S_{XY}/R_{XY} + 1.994^HAFPSA^{(2)} + 0.594(HA,HD)^{max/min}$
18	$\log L = 0.489 + 1.059^1\chi + 0.377^1IC + 0.063RPCS - 0.0032DPSPA^{(1)} + 13.739FNSA^{(3)}$
19	$\log L = -0.850 + 1.496^1\chi + 1.754^1BIC + 0.062PNSA^{(2)}$
20	$\log L = -1.234 + 1.541^1\chi + 0.057PNSA^{(2)} + 0.404\Phi$
21	$\log L = -1.293 + 0.534^3\chi + 0.039WNSA^{(1)} + 2.000FPSPA^{(2)} + 1.763^0IC + 28.440FNSA^{(3)}$
22	$\log L = -1.329 + 0.053WNSA^{(1)} + 0.288^3\chi + 0.684WPSA^{(3)} + 0.344^0CIC + 2.410S_{XY}/R_{XY}$
23	$\log L = -1.357 + 0.027PNSA^{(1)} + 3.235FPSPA^{(2)} - 3.735^HDFPSA^{(2)} + 87.785FPSPA^{(3)} + 0.0078\Delta H_f^0$
24	$\log L = 44.908 + 0.061\alpha + 2.556N^{occ.el.le}v/N_A - 7.314E_{exc}^{min}(H - C) - 1.154E_{state}^{min}(H)$
Protic Solvents – Alcohols	
25	$\log L = -5.093 + 1.639N^{HA} - 8.861q_A^{min} + 1.144\mu_{hyb} + 0.959E_{tot} + 0.011^HAPSA^{(2)}$
26	$\log L = 0.249 + 1.074^1\chi + 1.089(HA,HD)^{max/min} + 0.115HACA^{(1)} + 12.568FPSPA^{(3)} - 0.027^1CIC$
27	$\log L = -1.087 + 0.895^1\chi + 1.298(HA,HD)^{max/min} + 4.438^1SIC$
28	$\log L = 0.081 + 1.269^1\chi + 1.574(HA,HD)^{max/min} + 0.088HDCA - 0.018^0CIC$
29	$\log L = -0.946 + 1.118^1\chi - 0.165PPSA^{(3)} + 2.165N^{HD} + 4.133^0SIC$
30	$\log L = 3.106 + 1.108(HA,HD)^{max/min} + 0.447^0\chi + 5.737^HDFPSA^{(2)}$
31	$\log L = 3.918 + 1.239(HA,HD)^{max/min} + 0.434^0\chi - 0.0033PPSA^{(1)}$
32	$\log L = 2.276 + 1.316(HA,HD)^{max/min} - 0.106\Phi + 0.538^0\chi$
33	$\log L = 2.758 + 0.533^0\chi + 1.113(HA,HD)^{max/min} - 0.005DPSPA^{(2)}$
34	$\log L = 2.274 + 2.396(HA,HD)^{min} + 0.520^0\chi$
35	$\log L = 7.479 + 0.0065W + 1.475(HA,HD)^{max/min} - 0.010WPSA^{(1)} - 0.203PPSA^{(3)}$
36	$\log L = 2.238 + 1.088(HA,HD)^{max/min} + 0.522^0\chi$
37	$\log L = 2.264 + 1.142(HA,HD)^{max/min} + 0.541^0\chi + 1.138^HHDCA^{(2)}$
38	$\log L = 0.189 + 0.015S_M - 13.201FNSA^{(2)} + 0.030WPSA^{(1)}$
39	$\log L = 1.450 + 1.808(HA,HD)^{min} + 0.668^0\chi - 0.402FPSPA^{(2)}$
40	$\log L = -1.259 + 1.145^1\chi + 1.078(HA,HD)^{max/min} + 0.181HACA^{(1)} - 11.146q_A^{max} + 3.802^0SIC$
41	$\log L = -1.399 + 0.008G_b + 13.895P - 1.140T^E$
42	$\log L = 9.204 + 2.924\mu_{hyb} + 120.666R_C^{min}$
Polar Aprotic Solvents – Ethers	
43	$\log L = -0.022 + 0.056\alpha + 0.046WNSA^{(1)} + 0.193HDCA$
44	$\log L = -1.012 + 0.926^0\chi - 1.031FPSPA^{(2)} + 0.586^0IC + 0.185^HADS^{(2)}$
45	$\log L = 0.066 + 1.033^1\chi + 1.890^0BIC$
46	$\log L = -3.523 + 0.063\alpha + 0.025HASA + 0.866^0IC + 0.0052\Delta H_f^0 + 2.169n_A^{max}$
47	$\log L = -4.409 + 0.055\alpha + 1.218(HA,HD)^{max/min} - 20.036FNSA^{(3)} + 3.002\bar{P}_C + 0.967N^{occ.el.le}v/N_A$
Dipolar Aprotic Solvents	
48	$\log L = 1.497 + 1.179^1\chi + 33.285q_C^{max} - 0.327PPSA^{(3)} + 19.666FHACA$
49	$\log L = -0.912 + 1.151^1\chi + 1.469^0IC + 0.772(HA,HD)^{max/min} - 0.039S_{ZX} + 0.0043^HAPSA^{(2)}$
50	$\log L = 1.702 + 1.075^1\chi + 51.143q_C^{min} + 1.266(HA,HD)^{max/min} + 0.942^2SIC - 0.012PPSA^{(2)}$
51	$\log L = -0.045 + 1.142^1\chi + 0.434\mu - 0.0047DPSPA^{(1)} + 0.635^HDCPSA^{(2)}$
52	$\log L = 2.072 + 0.214N^{occ.el.le}v - 0.845J - 34.456FNSA^{(3)} + 60.910q_C^{min}$
53	$\log L = 9.483 + 0.0017G_p - 0.093PNSA^{(3)} - 1.942V_C^{min} - 0.391J + 0.667WPSA^{(3)}$
54	$\log L = 2.358 + 0.824^1\chi + 0.031WNSA^{(1)} - 1.245J$
55	$\log L = 6.627 + 0.0056G_b - 7.777V_H^{min} + 16.637q_C^{max}$
56	$\log L = -3.302 - 13.248q_{net}^{min} + 0.023PNSA^{(1)} + 0.938WPSA^{(3)} - 0.261HOMO - 0.759J$
57	$\log L = -3.543 + 0.0043G_b + 0.927E_{tot} + 1.517\mu_{hyb} + 5.241FHASA$
58	$\log L = -0.217 + 0.0027G_p - 27.326FNSA^{(3)} + 1.506\mu_{hyb}$
59	$\log L = 2.271 + 0.0045G_b + 1.753^0IC + 0.272HOMO + 9.280FHASA - 13.558q_A^{max}$
60	$\log L = -0.275 + 0.0056G_b + 0.924\nu_{TD}^h + 0.175\mu + 1.014HACA^{(2)} + 1.014FNSA^{(1)}$
61	$\log L = -0.568 - 0.218WNSA^{(2)} + 0.575^1\chi^v + 1.749\nu_{TD}^h + 1.147FNSA^{(1)}$

Table 2 (Continued)

eq	QSPR model
Dipolar Aprotic Solvents	
62	$\log L = -2.540 + 0.0061G_b - 6.797FPSA^{(3)} + 60.391^{HA}FCPSA^{(2)} + 4.350S_{YZ}/R_{YZ}$
63	$\log L = 2.834 - 0.113PNSA^{(3)} + 0.0017G_p - 0.867J + 0.041^{HD}PSA^{(2)} - 0.040^2SIC$
65	$\log L = -0.794 + 0.0027G_p + 1.529^0IC - 1.159WNSA^{(3)} - 1.015^{HA}FPSA^{(2)} - 0.206PPSA^{(3)}$
65	$\log L = -2.258 + 1.710^0IC + 0.969^1\chi + 1.622(HA,HD)^{max/min} - 0.077\Delta H_f^0/N_A + 0.0082PNSA^{(1)}$
66	$\log L = 0.412 + 0.053WNSA^{(1)} + 6.657P_p^2 + 0.049PPSA^{(3)}$
67	$\log L = 23.173 + 2.113HOMO + 0.103PPSA^{(3)}$
68	$\log L = 0.598 + 0.595\mu_{pch} + 2.816FHBSA + 0.707^1\chi - 0.0029DPSPA^{(1)}$
69	$\log L = 0.901 + 0.0092PNSA^{(1)} + 0.0041G_b + 84.302^{HA}FPSA^{(2)} - 1.014c_p^{int}/N_A$

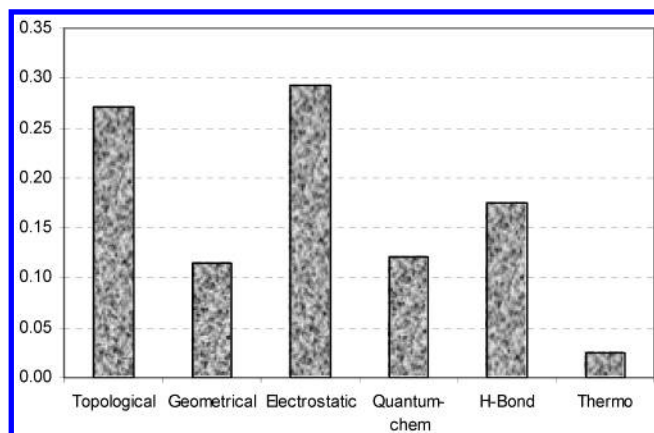


Figure 2. Relative distribution of different types of descriptors over all solvents.

account for the effective mass distribution in a molecule, providing a good measure of the dispersion forces in the bulk solvent. The topological family of descriptors⁷⁰ is represented by the average information content of order 2 (2IC) and the complementary information content of orders 0 and 1 (0CIC and 1CIC respectively). These descriptors, of different orders, account for both the constitutional and structural diversity of a molecule. The zeroth-order indices code chemical diversity realized as the presence of atoms of types other than C, H, and O. The information topological indices of higher orders also account for the branching and diversity of the molecular structure.

Inert Solvents (ii) – Aromatics and Chloroaromatics. Benzene and alkyl- and halogeno-benzenes exhibit similar solvent properties. The QSPR models for them presented in Table 1 are of good statistical quality, with R^2 and s^2 ranging from 0.926 and 0.52 for benzene to 0.974 and 0.37 for *m*-xylene, respectively. The Randic indices of different orders, or molecular polarizability in the case of bromobenzene, are frequently used molecular bulk descriptors. The number of occupied electronic levels ($N^{occ.el,lev}$) is another related descriptor because the higher the number of the electronic shells, the more polarizable the molecule.

Electrostatic interactions are reflected by contributions from the family of surface area descriptors such as the partial negative surface area (PNSA), the partial positive surface area (PPSA), or functions of them such as the fractional PPSA (FPSA) and the difference between PPSA and PNSA (DPSPA). Five of the eight models include topological descriptors such as the average information content (IC), bond information content (BIC), and complementary information content (CIC). The information content of different types and different orders reflects the diversity of atomic and structural constitution of organic molecules.

Two models (benzene and *m*-xylene) include the Kier–Hall flexibility index (Φ), a topological index based on structural properties which is considered as measuring the constraints on a molecule attaining infinite flexibility (like an endless chain of C sp^3 atoms). The flexibility index depends on (a) number of atoms, (b) the presence of rings, (c) branching, and (d) the presence of atoms with covalent radii smaller than those of C sp^3 .

Protic Solvents – Alcohols. Hydroxyl-containing solvents are represented by water, 16 aliphatic alcohols, 1 aromatic alcohol, and 1 diol (ethylene glycol). Solubilities in water and 1-octanol have been measured for the largest numbers of solutes: 177 and 173, respectively. Though only 14 data points were available, the model for ethylene glycol includes only two adjustable parameters. The models for these solvents are characterized by good to excellent statistical quality with R^2 varying from 0.878 for 2-ethyl-1-hexanol to 0.982 for benzyl alcohol.

As expected, hydrogen bond descriptors are of major importance in modeling the protic solvents. The HB descriptors include simple integer examples such as the counts of hydrogen acceptor or donor sites (N^{HA} , N^{HD}) together with the ratio of the maximal number of hydrogen bond donor or acceptor sites in a molecule to the corresponding minimal value, $(HA, HD)^{min/max}$, and advanced hydrogen bond descriptors expressed in the form of partial surface area. The latter include the hydrogen acceptor charged surface area (HACA) and the hydrogen donor charged surface area (HDCA).

Electrostatic interactions are also relevant to solvation in protic solvents. The first two terms of the multipole expansion¹ serve as descriptors in the QSPR models for water, 1-octanol, and ethylene glycol: the minimum and maximum atomic charge for all atom types (q_A^{min} and q_A^{max}) and the hybridization component of the molecular dipole (μ_{hyb}). The μ_{hyb} descriptor contributes positively to the solubility because dipole–dipole interactions are strong in alcohol media. The topographic electronic index, T^E , a more complex function of the atomic charges and molecular geometry, is calculated as the sum of the charge differences over all molecular bonds. Equation 41 for benzyl alcohol contains descriptor T^E , bearing the negative sign. This is possibly due to the fact that large values of T^E are possessed not only by compact, highly polar molecules but also by long-chain lower polarity ones. The charge-weighted (FPSA and PNSA) and the total molecular surface area weighted (WNSA) partial surface area appear in 6 out of the 18 equations pointing to the high importance of electrostatic interactions on solvating protic species, which are of high polarity.

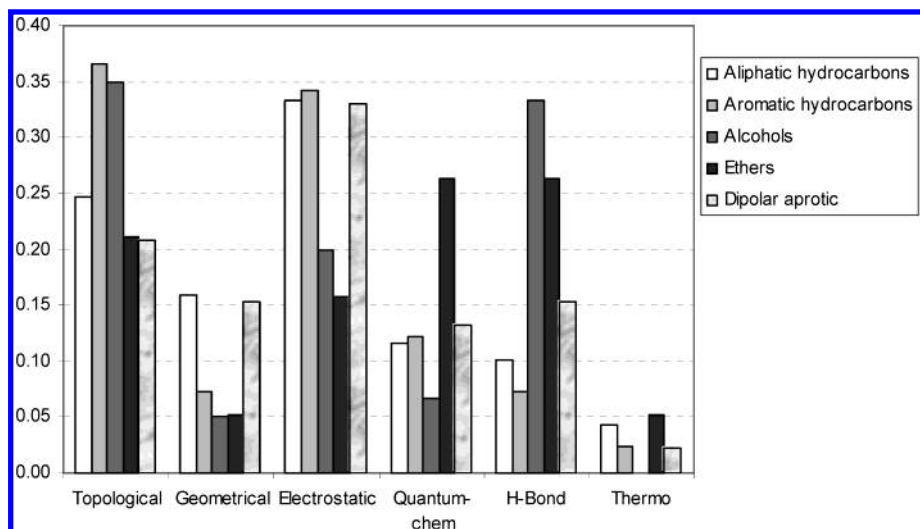


Figure 3. Relative distribution of different types of descriptors over particular solvent groups.

The cavity-forming features of the molecules are adequately described by topological indices, of which the Randic indices of zero- and first-orders occur in lines 26–40 of Table 2 and various information topological indices (*IC*, *CIC*, *SIC*) in lines 26–29 and 40. The coefficients of eqs 25–42 suggest that the cavity formation term is of less significance in describing solvating ability of protic solvents than the contributions from the hydrogen bond descriptors, such as (*HA*, *HD*)^{min}, *HACA*, and *HDCA*. This observation is in agreement with the Abraham approach.¹⁸

Polar Aprotic Solvents – Ethers. Three acyclic and two cyclic ethers (THF and 1,4-dioxane) comprise the group of five polar aprotic solvents, lines 43–47 of Tables 1 and 2. The statistical quality of these QSPR models varies from good for methyl *tert*-butyl ether ($R^2 = 0.944$, $s^2 = 0.342$) to excellent for dipropyl ether ($R^2 = 0.998$, $s^2 = 0.016$).

As solvents, ethers demonstrate good solvating properties due to their high polarity and Lewis basicity. The Lewis basicity is directly reflected by such descriptors as *HASA* (the hydrogen bond acceptor surface area) and (*HA*, *HD*)^{max/min}, both of which bear the positive sign in eqs 46 and 47. For hydrophilic solutes, a still more important parameter is their HB donor ability, that is, the ability to interact with the lone pairs of the ether oxygen. In the case of methyl *tert*-butyl ether, where the hydrogen bonding is sterically hindered, only cavity-formation terms contribute significantly to the property: eq 45 includes only two topological indices such as the Randic first-order index (χ) and the average bond information content of zero-order (0BIC).

Other important cavity-formation descriptors in eqs 43–47 are molecular polarizability, α , the information topological indices, and the number of occupied electronic levels ($N_{occ,el,lev}$). The electrostatic contribution to the solvation processes in ethers is described by such polar surface areas as *WNSA*⁽¹⁾, *FPSA*⁽²⁾, and *FNSA*⁽³⁾.

Dipolar Aprotic Solvents. The most comprehensive solvent data set is that of the dipolar aprotic species, which includes 22 diverse solvents. The best QSPR model is derived for solubilities in adiponitrile ($R^2 = 0.994$, $s^2 = 0.004$), in *N*-methyl-2-pyrrolidone ($R^2 = 0.994$, $s^2 = 0.059$), and in *N,N*-dimethylformamide ($R^2 = 0.990$, $s^2 = 0.065$). The correlation for 2-pyrrolidone is far poorer ($R^2 = 0.837$, $s^2 = 0.093$) and could be explained by an insufficient ability of

the existing CODESSA PRO descriptors to treat interactions with molecules as complex as 2-pyrrolidone: which is simultaneously a free nitrogen base, a hydrogen bond donor, and a dipolar aprotic species. The QSPR model is much better for the corresponding methyl derivative, *N*-methyl-2-pyrrolidone.

The dipolar aprotic nature of the species listed in Tables 1 and 2, lines 48–69, is reflected by the descriptors that characterize the solutes under study. The dipole–dipole and other electrostatic interactions seem to be the most significant intermolecular interactions. That applies to most of the QSPR models: atomic charges, dipole moments, and various charged surface area descriptors are those most frequently used in models 48–69. The molecular dipole moment or its components occur explicitly in five models, eqs 51, 57, 58, 60, and 68. The contribution of the monopole moment is reflected by different atomic charges, as reflected in six equations (48, 50, 52, 55, 56, and 59).

However, the cavity-formation contribution also turns out to be significant in the modeling of the solvating ability of dipolar aprotic solvents. The first-order Randic index, the Kier–Hall valence connectivity index, the Balaban index, and different information topological indices participate in 13 out of 22 models. The Balaban index, bearing the negative sign in eqs 52–54, 56, and 63, describes the negative effect of molecular “centricity” on solubility in dipolar aprotic solvents. The Balaban index takes its largest values in highly branched hydrocarbons, which are poorly soluble in the dipolar aprotic solvents studied.

The gravitational indices¹⁹ G_b and G_p (calculated over all bonds or all atom pairs) are further measures of molecular bulk properties or the ability to display dispersion interactions. Eight QSPR models (eqs 53, 55, 57–60, 62–64, and 69) contain G_b and G_p bearing a positive sign. Another measure of dispersion interactions is the energy of the highest occupied molecular orbital (HOMO). The hydrogen bonding term is also important: (*HA*, *HD*)^{max/min}, *HACA*, *HASA*, and *FHBSA* (the fractional hydrogen bond surface area) contribute to 8 out of the 22 models.

CONCLUSIONS

The QSPR solvation models, based on the CODESSA PRO theoretical descriptors, are fast and have reasonable

predictive power. The squared standard deviations or variances of the predictive models range from 0.004 for adiponitrile to 0.608 for ethanol, measured in logarithmic units of the Ostwald solubility coefficients. About 500 organic solutes were studied over the 69 solvents from 14 solutes for ethylene glycol to 226 for hexadecane. All the QSPR models are built on the assumption of the strong nonorthogonality of the underlying descriptors: the default cutoff for descriptor one-to-all-others intercorrelation is fixed as 0.5. The methodology developed is potentially useful for the study of a range of diverse phenomena, including fluid-phase equilibria and ADME evaluations.

EXPERIMENTAL SECTION

The two- and three-dimensional molecular structures of solutes were stored with their properties in an ISIS/Base database.⁷¹ 2D to 3D conversions and preoptimization were performed using Hyperchem 5.1.⁷² Final optimizations were performed using the AM1 parametrization⁷³ within the semiempirical quantum-chemical program CMOPAC (implemented in CODESSA PRO⁶⁰), which is based on MOPAC version 7.⁶¹ A gradient norm 0.01 kcal/Å was forced to calculate electronic, geometric, and energetic parameters for the isolated molecules. FORCE calculations were used to produce thermodynamic parameters. The structures were used in the CODESSA PRO program to calculate constitutional, topological, geometrical, thermodynamic, quantum chemical, and electrostatic descriptors.^{45,74}

The total number of CODESSA PRO descriptors computed for the given data set is 834. Some descriptors, which are clearly irrelevant to solubility, were deliberately excluded from the descriptor pool. To include the most appropriate ones, we formed special descriptor sublists, discarding such low-informative or inappropriate variables as constitutional descriptors and quantum chemical reactivity indices. The number of descriptors for a particular structure is dependent on the number of possible results for the invariant operators over the atomic and bond constituents. Correlations were produced with HM PRO (the Heuristic method for CODESSA PRO), which has an algorithm consisting of four major parts:

(1) The one-parameter descriptors selection. The selection is based on squared correlation coefficient, Fisher F -criteria, and Student t -criteria. Eliminating of the highly intercorrelated descriptors and descriptors with insignificant variance is realized.

(2) Pairwise selection. This selection is made on the basis of the squared correlation coefficient and Fisher F -criteria.

(3) Expanding/contracting stage. The unexpanded correlations in the correlation set are expanded by adding a previously unselected descriptor. The number of correlations added in such a manner can be limited by values of the squared intercorrelation coefficient, F -criteria (normalized or not), and standard error. The correlations with the maximum allowed number of descriptors (given parameter) are not expandable. The correlations with a minimum number of descriptors (i.e. two descriptors, because all one-parameter correlations are in the in-memory correlation set) are not contractible. The process begins from the correlations with largest values of the fitness function. The identical correlations are not put into in-memory correlation set (uniqueness

is constrained). This stage will be repeated until a stop event appears, which can be any/all of following:

The in-memory correlation set is overfilled (It can be conditionally unlimited when the correlation value of fitness function is less than minimal in the set, and it is not stored at all after overfilling. In this case, if correlation is inserted into the set, the correlation with the worst value of the fitness function is eliminated).

Maximum number of iterations achieved.

Time limit is reached.

The correlation set does not have any correlation to expand/contract (full search is finished).

(4) The output stage. The given number of the "best" correlations is printed out. The iterations for selecting for the printed correlations begin from the "best" correlations. For all correlations in a cycle, the full set of statistical parameters is calculated including intercorrelation of the descriptors (one to all others), cross-validated squared correlation coefficient, etc. The parameters of the method are defined by the set of the selection criteria. For any correlation, a full list of the predecessors in order of calculation can be printed and on the basis of best correlations with subsets of the descriptors until a one-parameter correlation will be printed.

The fitness function above is defined as follows

$$w = (r^2)^{x_1} F^{x_2} n^{x_3} (s^2)^{x_4} N^{x_5}$$

where r^2 is the squared correlation coefficient, F is the Fisher criteria, n is the number of the data points, s^2 is the standard error (nonnormalized), and N is the number of the descriptors in the correlation. The x_i is weight of the factors and is selected as {1, 1, 1, -1, -1}.

ACKNOWLEDGMENT

We thank Professor M. Karelson for very helpful discussions and Ms. Hongfang Yang and Ms. Katherine Kovalenko for help in MS preparation. One of the authors (U.M.) is grateful to the Estonian Science Foundation (Grant #4571) for the financial support.

Supporting Information Available: Experimental and calculated solubilities of solvent series (Table 3 (3-1–3-69)), explanation of the descriptors used in Table 2 (Table 4), and data references (for Table 3) (Table 5). This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- Reichardt, C. *Solvent and Solvent Effects in Organic Chemistry*; VCH: Weinheim, 1988.
- Quantitative Treatments of Solute/Solvent Interactions*; Polizer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994.
- Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1–18.
- Katritzky, A. R.; Fara, D. C.; Petrukhin, R.; Tatham, D. B.; Maran, U.; Lomaka, A.; Karelson, M. The Present Utility and Future Potential for Medicinal Chemistry of QSAR/QSPR with Whole Molecule Descriptors. *Top. Curr. Med. Chem.* **2002**, *2*, 1333–1356.
- Katritzky, A. R.; Petrukhin, R.; Tatham, D.; Basak, S.; Benfenati, E.; Karelson, M.; Maran, U. The Interpretation of Quantitative Structure–Property and –Activity Relationships. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 679–685.

- (6) Draper, N. R.; Smith, H. *Applied Regression Analysis*; Wiley: New York, 1981.
- (7) Bendel, R. B.; Afifi, A. A. Comparison of Stopping Rules in Forward "Stepwise" Regression. *J. Am. Stat. Assoc.* **1977**, *72*, 46.
- (8) Goldberg, D. E. *Genetic Algorithm in Search, Optimization, and Machine Learning*; Addison-Wesley: Reading, MA, 1989.
- (9) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure–Activity Relationships and Quantitative Structure–Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
- (10) Bohachevsky, I. O.; Johnson, M. E.; Stein, M. L. Generalized Simulated Annealing for Function Optimization. *Technometrics* **1986**, *28*, 209–217.
- (11) Sutter, J. M.; Dixon, S. L.; Jurs, P. C. Automated Descriptor Selection for Quantitative Structure–Activity Relationships Using Generalized Simulated Annealing. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 77–84.
- (12) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. Accurate Transferable Model for Water, *n*-Octanol, and *n*-Hexadecane Solvation Free Energies. *J. Phys. Chem. B* **2002**, *106*, 11009–11015.
- (13) Duffy, E. M.; Jorgensen, W. L. Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, *122*, 2878–2888.
- (14) Ishikawa, T.; Tsuboi, A.; Kolar, P.; Shen, J.-W. Solvent Selection for Pharmaceutical. *Fluid Phase Equil.* **2002**, *194*–*197*, 771–782.
- (15) McElroy, N. R.; Jurs, P. C. Prediction of Aqueous Solubility of Heteroatom-Containing Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1237–1247.
- (16) Ran, Y.; Yalkowsky, S. H. Prediction of Drug Solubilities by the General Solubility Equation (GSE). *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 354–357.
- (17) Acree, W. E.; Powell, J. R.; McHale, M. E. R.; Pandey, S.; Borders, T. L.; Campbell, S. W. Thermodynamics of Mobile Order Theory. *Research Trends in Physical Chemistry, Council of Scientific Research Integration, Trivandrum, India*; 1997; Vol. 6, pp 197–233.
- (18) Abraham, M. H. Physicochemical and Biological Processes. *Chem. Soc. Rev.* **1993**, 73–83.
- (19) 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.
- (20) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (21) Duffy, E. M.; Jorgensen, W. L. Predictions of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, *122*, 2878–2888.
- (22) Viswanadhan, V. N.; Ghose, A. K.; Singh, U. C.; Wendoloski, J. J. Prediction of Solvation Free Energies of Small Organic Molecules: Additive-Constitutive Models Based on Molecular Fingerprints and Atomic Constants. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 405–412.
- (23) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, 300–313.
- (24) Sadus, R. J. *Molecular Simulation of Fluids: Theory, Algorithms, and Object-Oriented*; New York: Elsevier: 1999.
- (25) Murray, J. S.; Abu-Awwad, F.; Politzer, P. Prediction of Aqueous Solvation Free Energies from Properties of Solute Molecular Surface Electrostatic Potential. *J. Phys. Chem. A* **1999**, *103*, 1853–1856.
- (26) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Reidel: Dordrecht, 1980; p 25.
- (27) Karelson, M.; Tamm, T.; Katritzky, A. R.; Cato, S. J.; Zerner, M. C. Application of Self-Consistent Reaction Field Method in Semiempirical Quantum Chemical Calculations. *Tetrahedron Comput. Methodol.* **1989**, *2*, 295–304.
- (28) Karelson, M. M. Ab initio Self-Consistent Reaction Field Calculations. I. Introduction. *Org. React.* **1980**, *17*, 357–365. Karelson, M. M. Ab initio Self-Consistent Reaction Field Calculations. II. Dipolar Molecules. *Org. React.* **1980**, *17*, 366–383. Karelson, M. M. Ab initio Self-Consistent Reaction Field Calculations. III. Molecular Ions. *Org. React.* **1983**, *20*, 127–140.
- (29) Tomasi, J.; Persico, M. Molecular Interactions in Solution: an Overview of Methods Based on Continuous Distribution of the Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.
- (30) Klamt, A.; Schuurmann, G. COSMO: a New Approach to Dielectric Screening in Solvents with Explicit Expression for the Screening Energy and its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.
- (31) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Extension of the Platform of Applicability of the SM5.42R Universal Solvation Model. *Theor. Chem. Acc.* **1999**, *103*, 9–63.
- (32) Torens, F. Universal Organic Solvent–Water Partition Coefficient. Model *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 236–240.
- (33) Abboud, J. L.; Taft, R. W. Regarding a Generalized Scale of Solvent Polarities. *J. Am. Chem. Soc.* **1977**, *99*, 8325–8327.
- (34) Taft, R. W.; Abraham, M. H.; Famini, G. R.; Doherty, R. M.; Abboud, J. L. M.; Kamlet, M. J. Solubility Properties in Polymers and Biological Media. 5. An Analysis of the Physicochemical Properties which Influence Octanol Water Partition-Coefficients of Aliphatic and Aromatic Solutes. *J. Pharm. Sci.* **1985**, *74*, 807–814.
- (35) Abraham, M. H.; Andonian-Hafvtan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors that Influence the Solubility of Gases and Vapors in Water at 298 K, and a New Method for its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.
- (36) Abraham, M. H.; Le, J.; Acree, W. E., Jr.; Carr, P. W. Solubility of Gases and Vapors in Propan-1-ol at 298 K. *J. Phys. Org. Chem.* **1999**, *12*, 675–680.
- (37) Abraham, M. H.; Le, J.; Acree, W. E. The Solvation Properties of the Aliphatic Alcohols. *Collect. Czech. Chem. Commun.* **1999**, *64*, 1748–1760.
- (38) Abraham, M. H.; Platts, J. A.; Hersey, A.; Leo, A. J.; Taft, R. W. Correlation and Estimation of Gas-Chloroform and Water-Chloroform Partition Coefficients by a Linear Free Energy Relationship Method. *J. Pharm. Sci.* **1999**, *88*, 670–679.
- (39) Abraham, M. H.; Whiting, G. S.; Doherty, R. M.; Shuely, W. J. Hydrogen Bonding. Part 14. The Characterization of Some *n*-Substituted Amides as Solvents: Comparison with Gas–Liquid Chromatography Stationary Phases. *J. Chem. Soc., Perkin Trans. 2* **1990**, 1851–1857.
- (40) Abraham, M. H.; Andonian-Haftvan, J.; Osei-Owusu, J.; Sakellariou, P.; Urieta, J. S.; López, M. C.; Fuchs, R. Hydrogen Bonding. Part 25. The Solvation Properties of Methylene Iodide. *J. Chem. Soc., Perkin Trans. 2* **1993**, 299–304.
- (41) (a) Huyskens P. L. Effect of the Fluctuations of the Potential Energy Curves Ruling the Motions of Atoms on the Thermodynamics of Liquids. *J. Mol. Struct.* **1992**, *270*, 197–203. (b) Huyskens P. L.; Haulait-Pirson M. C. Influence of H–Bond Chains in Solvents on the Solubility of Inert Substances. A New Quantitative Approach. *J. Mol. Liq.* **1985**, *31*(3), 153–176.
- (42) Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R. O.; Tatham, D. B.; Maran, U.; Lomaka, A.; Acree, W. E. A General Treatment of Solubility. 2. QSPR Prediction of Free Energies of Solvation of Specified Solutes in Ranges of Solvents. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1806–1814.
- (43) Katritzky, A. R.; Lobanov, V.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279–287.
- (44) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indices in Organic Chemistry. *Russ. Chem. Rev. (Uspekhi Khimii)* **1988**, *57*, 337–366.
- (45) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027–1043.
- (46) Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley: New York, 2000.
- (47) Katritzky, A. R.; Tatham, D. B.; Maran, U. The Correlation of the Solubilities of Gases and Vapors in Methanol and Ethanol with Their Molecular Structures. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 358–363.
- (48) 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.
- (49) Tetko, I. V.; Tanchuk, V. Yu.; Kasheva, T. N.; Estimation of Aqueous Solubility of Chemical Compounds Using E-State Indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1488–1493.
- (50) Huuskonen J.; Salo, M.; Taskinen, J. Aqueous Solubility Prediction of Drugs Based on Molecular Topology and Neural Network Modeling. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 450–456.
- (51) Huuskonen, J. Estimation of Aqueous Solubility for a Diverse Set of Organic Compounds Based on Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 773–777.
- (52) Klopman, G.; Zhu, H.; Estimation of the Aqueous Solubility of Organic Molecules by the Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 439–445.
- (53) Klopman, G.; Zhu, H. Estimation of the Aqueous Solubility of Organic Molecules by the Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 439–455.
- (54) Viswanadhan, V. N.; Ghose, A. K.; Wendoloski, J. J. Estimating Aqueous Solvation and Lipophilicity of Small Organic Molecules: A Comparative Overview of Atom/Group Contribution Methods. *Perspect. Drug. Discov.* **2000**, *19*, 85–98.
- (55) Famini, G. R.; Wilson, L. Y. Using Theoretical Descriptors in Linear Solvation Energy Relationships. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 213–242.
- (56) Lowrey, A. H.; Cramer, C. J.; Urban, J. J.; Famini, G. R. Quantum Chemical Descriptors for Linear Solvation Energy Relationships. *Comput. Chem.* **1995**, *19*, 209–215.

- (57) Abraham, M. H.; Grellier, P. L.; McGill, R. A. Determination of Olive Oil-Gas and Hexadecane-Gas Partition-Coefficients, and Calculation of the Corresponding Olive Oil-Water and Hexadecane-Water Partition-Coefficients. *J. Chem. Soc., Perkin Trans. 2* **1987**, 797–803.
- (58) Acree, W. E.; Abraham, M. E. Solubility Predictions for Crystalline Polycyclic Aromatic Hydrocarbons (PAHs) Dissolved in Organic Solvents Based Upon the Abraham General Solvation Model. *Fluid Phase Equil.* **2002**, 201, 245–258.
- (59) *Handbook of Physical Properties of Organic Chemicals*; Howard, P. H., Meylan, W. M., Eds.; Lewis Publishers: Boca Raton, FL, 1997.
- (60) <http://www.codessa-pro.com>.
- (61) Stewart, J. J. P. *MOPAC Program Package*. QCPE 1989, No 455.
- (62) Popovych, O. *Nonaqueous Solution Chemistry*; Wiley: New York, 1981.
- (63) Karelson, M. Quantum Chemical Treatment of Molecules in Condensed Disordered Media. *Adv. Quantum Chem.* **1997**, 28, 141–157.
- (64) Cronce, D. T.; Famini, G. R.; De Soto, J. A.; Wilson, L. Y. Using Theoretical Descriptors in Quantitative Structure–Property Relationships: Some Distribution Equilibria. *J. Chem. Soc., Perkin Trans. 2* **1998**, 1293–1301.
- (65) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. Thermodynamics of Solute Transfer From water to Hexadecane. *J. Chem. Soc., Perkin Trans. 2* **1990**, 291–300.
- (66) Abraham, M. H.; Whiting, G. S.; Carr, P. W.; Ouyang, H. Hydrogen Bonding. Part 45. The Solubility of Gases and Vapours in Methanol at 298 K: An LFER Analysis. *J. Chem. Soc., Perkin Trans. 2* **1998**, 1385–1390.
- (67) Abraham, M. H.; Whiting, G. S.; Shuely, W. J.; Doherty, R. M. The Solubility of Gases and Vapours in Ethanol – the Connection Between Gaseous Solubility and Water–Solvent Partition. *Can. J. Chem.* **1998**, 76, 703–709.
- (68) Abraham, M. H.; Le, J.; Acree, W. E.; Carr, P. W.; Dallas, A. J. The Solubility of Gases and Vapours in Dry Octan-1-ol at 298 K. *Chemosphere* **2001**, 44, 855–863.
- (69) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, 97, 6609–6615.
- (70) Roy, A. B.; Basak, S. C.; Harris, D. K.; Magnuson, V. R. Neighborhood Complexities and Symmetry of Chemical Graphs and Their Biological Applications. In *Mathematical Modeling in Science and Technology*; Avula, X. J. R., Kalman, R. E., Liapis, A. I., Rodin, E. Y., Eds.; Pergamon Press: New York, 1984; pp 745–750.
- (71) ISIS/Base, MDL Information Systems, Inc., San Leandro, CA (www.mdli.com).
- (72) HyperChem 5.02, HyperCube, 1997.
- (73) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, 107, 3902–3909.
- (74) Stanton, D. T.; Jurs, P. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure–Property Relationship Studies. *Anal. Chem.* **1990**, 62, 2323–2329.

CI034120C