A General Treatment of Solubility. 2. QSPR Prediction of Free Energies of Solvation of Specified Solutes in Ranges 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

As part of our general QSPR treatment of solubility (started in the preceding paper), we now present quantitative relationships between solvent structures and the solvation free energies of individual solutes. Solvation free energies of 80 diverse organic solutes are each modeled in a range from 15 to 82 solvents using our CODESSA PRO software. Significant correlations (in terms of squared correlation coefficient) are found for all the 80 solutes: the best fit is obtained for *n*-propylamine ($R^2 = 0.996$); the lowest R^2 corresponds to toluene (0.604).

INTRODUCTION

The structures of both solute and solvent determine the interactions relevant to the solubility process. Consequently, a quantitative treatment of solute−solvent interactions may be expressed through a general formula as follows

$$P_{SOLUBILITY} = C_0 + \sum C_{el} D_{el-solvent} D_{el-solute} + \sum C_{disp} D_{disp-solvent} D_{disp-solute} + \sum C_{cav} D_{cav-solvent} D_{cav-solute} + \sum C_{HB} D_{HB-solvent} D_{HB-solute}$$

where (i) C_0, C_{el}, C_{disp}, C_{cav}, and C_{HB} are the intercept and the general coefficients for the electrostatic interaction, dispersion interaction, cavity formation, and solute−solvent hydrogen bonding terms, respectively; (ii) $D_{el/disp/cav/HB-solvent}$ are appropriate descriptors describing the properties of solvents; and (iii) $D_{el/disp/cav/HB-solute}$ are descriptors reflecting the properties of solutes. The summations indicate that each term can include more than one descriptor accounting for the same type of interaction. In a series for which the solute is constant, the solute descriptors can be combined into the corresponding coefficients, thus the solubility relationship depends solely on the structural features of the solvent. Conversely, when dealing with the solubility of different solutes in a single solvent, the solvent descriptors can be combined into the corresponding coefficients, and solubility is then determined only by the structural characteristics of the solutes.

Most quantitative treatments of solubility have expressed the structural properties of solutes in a series for a single solvent.[1,2] Our earlier work on gas solubilities in water,[3] in methanol, and in ethanol,[4] and the previous paper in the present series,[5] all describe studies in this direction. Quantitative treatment of solubility addressed by varying the structure of the solvent is less common. We now overview available publications for series of solubilities in which the solute is kept constant.

For these series, two experimentally based methods have been used to correlate and predict solubilities: *linear solvation energy relationships* (LSER) and *mobile order theory* (MOT). The LSER method is based on multilinear regression (MLR) analysis of the solubilities of solutes in different solvents and has gained increasing attention during the past decades. The method was originally developed by Kamlet and Taft[6,7] and further refined and applied by Abraham and co-workers[8] who have applied it to numerous solutes. These studies include the solubilities of anthracene,[9] phenanthrene,[9] *trans*-stilbene,[10] hexachlorobenzene,[9] ferrocene,[11] fullerene C_{60},[12] diuron,[13] and monuron.[13] The LSER MLR model utilizes several characteristics that account for 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 solid basis to discuss the solute−solvent interactions and also the ability to rank each type of interaction for each solute−solvent pair. A limitation is that the characteristics (descriptors) used in the LSER model originate from experimental measurements;

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

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

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

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

10.1021/ci034122x CCC: $25.00
Published on Web 10/01/2003

these are often unavailable or incomplete when working with diverse compounds within large databases.

The MOT approach[14] has been used extensively by Acree and co-workers to predict mole fraction solubilities of various solutes in nonelectrolyte solvents. The solutes studied include anthracene,[15] phenanthrene,[9] pyrene,[16] acenaphthene,[17] fluoranthene,[18] *trans*-stilbene,[19,20] benzil,[21] thianthrene,[22] thioxanthen-9-one,[23] diphenyl sulfone,[24] hexachlorobenzene,[9] ferrocene,[25] 4-nitroaniline and 4-nitro-*N*,*N*-dimethylaniline,[26] and diuron[27] and monuron.[28] The MOT is based on a thermodynamic treatment of the liquid state that includes terms for describing 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. Partners of hydrogen bonds are not preserved with time but rather change continuously. Such treatment leads to an equilibrium consideration involving the fractions of time during which an amphiphilic proton belongs respectively to a bonded and nonbonded state. This differs from the 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 solubility of anthracene and other polyacenes in different solvents was the subject of several studies[9,15] including extensive experimental and theoretical work by Acree and co-workers (see ref 14 and references herein). Recently, Acree and Abraham[9] reported a theoretical study on the solubility of anthracene, phenathrene, and hexachlorobenzene using LSER, MOT, and the UNIFAC group contribution approach. LSER five-parameter general solubility equations for the solubility of anthracene (in terms of Ostwald solubility coefficients in logarithmic units) in 29 solvents gave average absolute deviations of 1.7% for the equation relating partition coefficients between water and organic solvent (log *P*) and 1.07% for partition coefficients between the gas phase and a given solvent (log *L*). When the MOT equations were applied, the prediction results had an average percentage error of 3.2%. Two versions of the UNIFAC model resulted in predictions with 2.4 and 1.8% errors. The solubility of phenathrene was predicted with accuracy similar to that for anthracene[9] by applying LSER equations to derive two correlations for its solubility in 23 solvents with average absolute deviations of 1.2% and 2.0% for log*P* and log*L*, respectively. The respective MOT application to phenathrene results in a 2.04% prediction error. The solubility of hexachlorobenzene was studied in a range of 20 different solvents applying both a LSER equation and MOT.[9] The former method provided the better results, with the average absolute deviation of 1.9% of the Ostwald solubility coefficient logarithmic scale.

Extended studies of *trans*-stilbene solubilities in a range of organic solvents conducted by Abraham, Acree, and co-workers[10,19,20] resulted in good predictions in 17 nonaqueous solvents with the average absolute deviation of 1.2%.[10] More recently, Acree et al.[20] applied the MOT and reported a small average percentage error (0.9%) for *trans*-stilbene solubilities in 34 organic solvents.

Studies of the solubility of ferrocene (based on LSER methodology) resulted in a 1.6% average error for logarithmic Ostwald solubility coefficients for 19 solvents.[11] A comprehensive experimental study of ferrocene solubilities performed by Acree and co-workers[25] correlated ferrocene solubilities in 42 organic solvents using MOT with an average absolute deviation of 3.7%.

Solubilities of the pesticides *N*′*N*′-dimethyl-*N*-(3,4-dichlorophenyl)urea (diuron) and *N*′*N*′-dimethyl-*N*-(4-chlorophenyl)urea (monuron) were also investigated by Abraham, Acree, and co-workers.[13,27,28] A prediction based upon MOT for the solubility of diuron in 28 nonalcoholic solvents provided reasonable estimates with an average absolute deviation of 2.3%. The same authors,[13] using the LSER solvation equation, correlated the solubilities of diuron in 22 solvents with an average percentage error of 1.1%. For monuron solubilities in 25 solvents, the error was 1.1%; the corresponding result using MOT for 21 solvents was 2.4%.[28]

As already mentioned, MOT methodology has been used extensively in the prediction of solubility. The solubility of pyrene was predicted with a 2.4% average absolute deviation for a set of 30 organic solvents.[16,18] For acenaphthene in 29 solvents, the average absolute deviation between the predicted and observed values of the logarithmic Ostwald coefficients[17] was 1.8%. An average error of 2.2% was found for fluoranthene solubilities in 42 organic solvents, providing acetonitrile (a strong outlier) was omitted.[18] For 1,2-diphenylethane-1,2-dione (benzil), the discrepancy between experimental and MOT predicted solubility values for 30 solvents is 1.8% of the average absolute deviation.[21] The solubility of thianthrene was correlated in 20 organic solvents with an average absolute deviation of 2.5%.[22] The situation with thioxanthen-9-one is akin to that of the case of thianthrene discussed above. Acree et al.[23] determined experimental solubilities of thioxanthen-9-one in 35 different organic solvents and correlated 26 of them by MOT with an error of 4.3%.

Abraham et al. analyzed the solubility of fullerene in 20 solvents and applied an LSER equation[8] with an average absolute deviation of 1.4%.[12]

The application of purely theoretical molecular descriptors has seldom been used in studies of solubility series with the solute constant but has found application in a study of the solubility of fullerenes. Sivaraman et al.[29] used only topological and constitutional descriptors in modeling the solubility of fullerene. Although the predictions for the relatively small subsets are good, their treatment utilizes (i) valence connectivity indices of different orders which are highly intercorrelated and (ii) an indicator variable containing several hidden parameters. Jurs and co-workers[30] have also predicted fullerene solubility using MLR and feed-forward computational neural networks (CNN). Their final CNN architecture 9-3-1 resulted in a model that consisted of topological, geometric, and electronic descriptors, which tends to agree with basic solvation principles. Their model has root-mean-square errors of 0.255, 0.253, and 0.346 log solubility units for the training, cross-validation, and external prediction set, respectively.

In a very recent publication, Shang et al.[31] used ab initio quantum chemical calculations to collect a set of theoretical descriptors for 78 pure solvents. Following this, correlations

Table 1. Solutes and Corresponding Statistics for QSPR Models in a Series of Solvents[a]

no.	name of the solute	R^2	R^2_{cv}	s^2	F	n	N_D	no.	name of the solute	R^2	R^2_{cv}	s^2	F	n	N_D
							Hydrocarbons: Aliphatic								
1	*n*-pentane	0.881	0.839	0.019	37	31	5	**10**	2,5-dimethylhexane	0.923	0.858	0.020	55	29	5
2	*n*-hexane	0.870	0.806	0.031	36	33	5	**11**	ethylcyclohexane	0.946	0.921	0.015	85	30	5
3	cyclohexane	0.912	0.867	0.018	58	34	5	**12**	*n*-nonane	0.881	0.804	0.042	34	29	5
4	2-methylpentane	0.912	0.855	0.014	50	30	5	**13**	1-hexene	0.864	0.760	0.026	23	15	3
5	*n*-heptane	0.883	0.826	0.034	42	34	5	**14**	isoprene	0.928	0.864	0.005	47	15	3
6	2,4-dimethylpentane	0.909	0.839	0.018	46	29	5	**15**	dichloromethane	0.826	0.680	0.021	22	29	5
7	methylcyclohexane	0.925	0.889	0.016	52	22	4	**16**	chloroform	0.777	0.549	0.047	17	30	5
8	*n*-octane	0.884	0.841	0.037	41	33	5	**17**	carbon tetrachloride	0.855	0.784	0.010	28	24	4
9	2,3,4-trimethylpentane	0.942	0.905	0.014	75	29	5	**18**	1,2-dichloroethane	0.820	0.656	0.026	25	14	2
							Hydrocarbons: Aromatic								
19	benzene	0.773	0.677	0.011	17	31	5	**30**	*trans*-stilbene	0.936	0.909	0.018	134	52	5
20	toluene	0.604	0.355	0.025	9	43	6	**31**	benzil[b]	0.911	0.857	0.030	63	37	5
21	ethylbenzene	0.876	0.818	0.014	28	16	3	**32**	thianthrene	0.926	0.849	0.004	72	28	4
22	*o*-cresol	0.956	0.905	0.017	94	17	3	**33**	thioxanthen-9-one	0.943	0.920	0.010	123	35	4
23	*p*-cresol	0.955	0.906	0.021	85	16	3	**34**	diphenyl sulfone	0.977	0.966	0.014	294	40	5
24	naphthalene	0.901	0.838	0.008	37	16	3	**35**	chlorobenzene	0.734	0.533	0.015	10	20	4
25	anthracene	0.823	0.782	0.045	70	82	5	**36**	hexachlorobenzene	0.898	0.866	0.014	67	44	5
26	phenanthrene	0.916	0.868	0.026	96	50	5	**37**	4-nitropyridine *N*-oxide	0.960	0.863	0.047	190	37	4
27	pyrene	0.852	0.808	0.042	77	73	5	**38**	methyl 4-hydroxybenzoate	0.923	0.863	0.022	44	15	3
28	acenaphthene	0.913	0.869	0.009	80	44	5	**39**	ferrocene	0.941	0.917	0.006	160	45	4
29	fluoranthene	0.908	0.849	0.031	83	48	5	**40**	fullerene	0.929	0.911	0.112	136	58	5
							Alcohols								
41	methanol	0.917	0.881	0.076	72	41	5	**47**	1-hexanol	0.906	0.862	0.033	39	26	5
42	ethanol	0.926	0.910	0.044	96	45	5	**48**	1-heptanol	0.976	0.958	0.007	200	19	3
43	1-propanol	0.927	0.891	0.038	79	37	5	**49**	2-methyl-1-propanol	0.976	0.961	0.014	152	15	3
44	2-propanol	0.894	0.816	0.061	34	21	4	**50**	2-methyl-2-propanol	0.960	0.930	0.026	95	16	3
45	1-butanol	0.925	0.881	0.036	69	34	5	**51**	phenol	0.972	0.946	0.012	161	18	3
46	1-pentanol	0.954	0.921	0.017	97	18	3								
							Organic Bases								
52	ethylamine	0.974	0.912	0.005	152	16	3	**57**	4-nitro-*N*,*N*-dimethylaniline	0.953	0.937	0.016	166	38	4
53	*n*-propylamine	0.996	0.993	0.0006	1013	15	3	**58**	monuron[b]	0.973	0.963	0.034	259	42	5
54	*n*-butylamine	0.978	0.970	0.004	191	17	3	**59**	diuron[b]	0.951	0.933	0.065	167	49	5
55	aniline	0.956	0.927	0.032	81	20	4	**60**	piroxicam[b]	0.749	0.567	0.368	13	22	4
56	4-nitroaniline	0.806	0.726	0.080	32	45	5								
							Organic Acids								
61	benzoic acid	0.889	0.831	0.042	56	41	5	**65**	ibuprofen[b]	0.735	0.549	0.343	13	24	4
62	2-hydroxybenzoic acid	0.892	0.839	0.087	54	31	4	**66**	diclofenac[b]	0.765	0.579	0.264	15	23	4
63	4-hydroxybenzoic acid	0.936	0.862	0.094	77	32	5	**67**	haloperidol[b]	0.934	0.815	0.078	61	17	3
64	4-aminobenzoic acid	0.921	0.864	0.190	55	24	4	**68**	paracetamol[b]	0.901	0.828	0.163	55	29	4
							Dipolar Aprotic Species								
69	methyl acetate	0.979	0.967	0.005	238	19	3	**75**	acetonitrile	0.935	0.870	0.047	48	22	5
70	ethyl acetate	0.947	0.924	0.010	103	35	5	**76**	acetone	0.951	0.926	0.012	117	36	5
71	propyl acetate	0.987	0.971	0.004	388	19	3	**77**	1,4-dioxane	0.946	0.909	0.010	95	33	5
72	butyl acetate	0.982	0.966	0.004	227	22	4	**78**	2-butanone	0.913	0.871	0.017	65	37	5
73	pentyl acetate	0.986	0.973	0.003	356	19	3	**79**	2-hexanone	0.993	0.990	0.001	515	15	3
74	methyl pentanoate	0.992	0.989	0.002	565	17	3	**80**	nitromethane	0.931	0.871	0.034	60	23	4

[a] R^2 − squared correlation coefficient, R^2_{cv} − cross-validated squared correlation coefficient, s^2 − squared standard deviation, F − Fisher criterion, n − number of points in data set, N_D − number of descriptors in QSPR model. [b] IUPAC nomenclature: benzil − 1,2-diphenyl-ethane-1,2-dione; monuron − *N*′*N*′-dimethyl-*N*-(4-chlorophenyl)urea; diuron − *N*′*N*′-dimethyl-*N*-(3,4-dichlorophenyl) urea; piroxicam − 4-hydroxy-2-methyl-*N*-(2-pyridyl)-2H-1,2-benzothiazine-3-carboxamide 1,1-dioxide; ibuprofen − α-methyl-4-(2-methylpropyl)- benzeneacetic acid; diclofenac − 2-(2,6-dichloroanilino)phenylacetic acid; haloperidol − 4-(4-hydroxy-4′-chloro-4-phenylpiperidino)-4′-fluorobutyrophenone; paracetamol − 4-hydroxy-acetanilide, *N*-(4-hydroxyphenyl)-acetamide.

were established for the solubilities of naphthalene, phenanthrene, anthracene, biphenyl, acenaphthene, hexachlorobenzene, benzyl, thioxanthene-9-one, diphenyl sulfone, and diuron. The quality of models varies from 0.861 to 0.931 of R^2 value.

In the present paper, we use the same database as in our preceding paper,[5] but now consider series in which the solute is constant, and thus the solubility variations are determined by the solvent structure. We apply the method of forward selection of descriptor scales to form QSPR models and analyze the descriptor content of the models from the point of view of solubility and solute−solvent interactions.

DATA AND METHODOLOGY

The general arrangement of the solubility data has already been described in detail in our preceding article.[5] For the current study, we selected 80 solutes, choosing only those that have reliable solubility data for at least 15 solvents. The solutes selected are listed in Table 1 along with the statistical parameters of the corresponding QSPR models.

The computational methodology applied to the current study coincides in general with that used in the preceding article.[5] Molecular structures of the solvents were drawn and optimized in the same fashion, and a total of 890 theoretical descriptors were calculated using CODESSA PRO software.

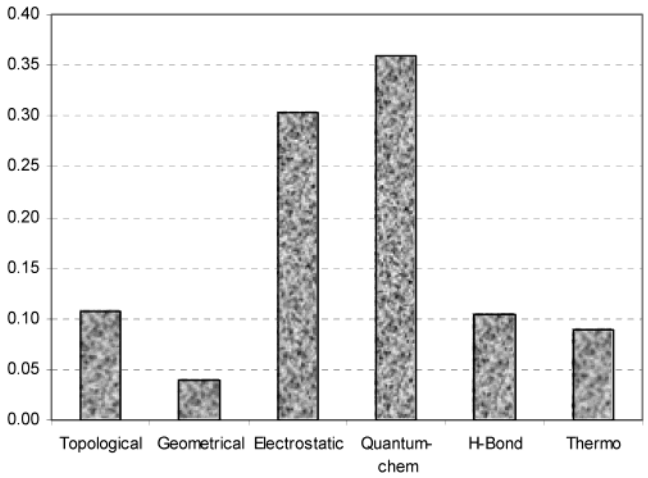

Figure 1. Relative distribution of different types of descriptors over all solutes.

The theoretical considerations of the interplay between CODESSA descriptors and different components of solvation free energy discussed in our preceding article are of significant relevance to the current study.

RESULTS

The correlation results for all of the 80 solutes are listed in Tables 1 and 2. Each entry of Table 1 provides the statistical characteristics of a QSPR model including the chemical name of the solute the squared correlation coefficient (R^2), the cross-validated squared correlation coefficient (R^2_{cv}), the variance or squared standard deviation, (s^2), Fisher criterion value (F), the number of experimental points in data set (n), and the number of descriptors in QSPR model (N_D). All the solutes are tentatively partitioned into six main classes: (i) "aliphatic hydrocarbons" comprising 12 alkanes, 2 alkenes, and 4 chloroalkanes; (ii) "aromatic hydrocarbons", 22, including one chlorocompound; (iii) "saturated alcohols", 11, including phenol; (iv) "organic bases", 9; (v) "organic acids", 8; (vi) "dipolar aprotic species", 12, including 6 esters, 3 ketones, 1 nitrile, 1 nitro compound, and dioxane.

Table 2 displays the quantitative−structure activity relationship equations deduced for all the 80 solutes. The equations are written in a linear notation; the key to the independent variables (descriptors) is given in the Discussion section. The number of descriptors involved in each multilinear regression analysis is rather small, varying from 6 for toluene to 2 for 1,2-dichloroethane. Analysis of the overall frequency of the descriptors, with respect to all solutes, is displayed in Figure 1 and with respect to particular solvent groups in Figure 2.

Tables 3-1−3-80 of the Supporting Information is a collection of all the numerical data for the experimental and predicted solubilities of each of 80 solutes. Each of these tables contains a unique in-house ID for each solvent, its chemical name, experimental solubilities of the solute, and predicted solubilities of this solute as well as original literature references to each solubility value.

DISCUSSION

To compare our current results with those discussed in our preceding paper devoted to the treatment of solvents,[5] it is appropriate to preface this discussion with a general view of the descriptors that occurred in the QSPR equations for solutes. Again, as in the case of solvents (part 1 of this series), electrostatic descriptors contribute significantly to the QSPR equations for solutes, Figure 1, but, in contrast to part 1, the histogram also demonstrates a high rating of quantum chemical descriptors and a relatively small contribution from topological indices and hydrogen bond descriptors. Within the current study, we relate the molecular structure of solvents to the partition characteristics of solutes, and thus the solvents molecular features are of great importance. As the solvent is to be considered as the bulk medium, the shape and volume characteristics of individual molecules are less important. In the case of the solute molecular structures studied in part 1 of the present series of papers, the contribution of topological and geometrical descriptors is higher because the solutes 3D characteristics are crucial to the formation of cavities in the bulk solvent. For the solvents studied in the present paper, the most important contributions come from the electrostatic and dispersion forces, which are adequately reflected by various quantum chemical descriptors.

In general, the statistical quality of the QSPR models for the solutes appear to be slightly inferior to those for the

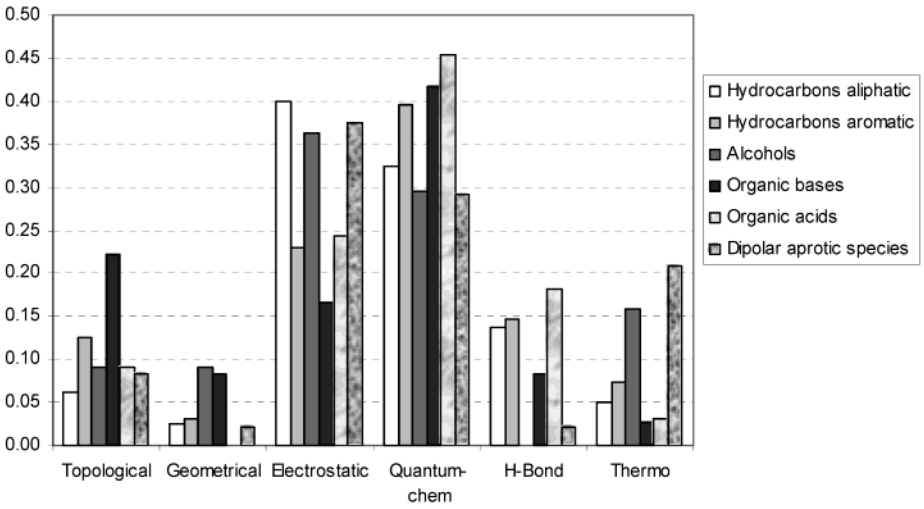

Figure 2. Relative distribution of different types of descriptors over particular solute groups.

Table 2. QSPR Models for Solubility of 80 Solutes in a Series of Solvents

no.	QSPR model
	Hydrocarbons: Aliphatic
1	$\log L = -10.531 - 1.975^{HA}FPSA^{(2)} - 17.110P^{max}_{\sigma-\pi} + 2.491\Delta E^{max}_{exc}(H-C) - 0.068HDCA + 0.0035PNSA^{(2)}$
2	$\log L = -18.777 - 2.195^{HA}FPSA^{(2)} - 0.836q^{max}_{net} + 8.140P^{min}_{H} + 0.0029PNSA^{(2)}$
3	$\log L = -5.498 - 1.578^{HA}FPSA^{(2)} + 1.187FNSA^{(2)} - 0.487q^{max}_{net} - 4.950FHBCA + 1.673\Delta E^{min}_{exc}(H-C)$
4	$\log L = -21.632 - 2.977\mu^2/MW - 0.550q^{max}_{net} + 2.280V^{min}_{H} + 4.238\Delta E^{max}_{exc}(H-C) - 0.016^2SIC$
5	$\log L = -2.748 - 0.0068^{HA}PSA^{(2)} - 0.070RPCS + 6.913P^{min}_{H} - 2.352^{HD}FPSA^{(2)} + 0.023PNSA^{(2)}$
6	$\log L = 0.971 - 3.283\mu^2/MW - 16.504P^{max}_{\sigma-\pi} - 6.650q^{min}_{C} + 2.876V^{min}_{H} - 0.054^1\chi$
7	$\log L = 5.986 - 1.878N^{occ.el.lev}/N_A - 0.045RNCS - 0.010DPSA^{(3)} - 7.932FHDCA$
8	$\log L = 1.360 - 0.982q^{max}_{net} + 2.736FPSA^{(1)} + 0.166WNSA^{(3)} - 19.700FHDCA + 26.322N^{max}_{C}$
9	$\log L = -3.533 - 3.823\mu^2/MW - 18.376P^{max}_{\sigma-\pi} - 7.023q^{min}_{C} + 0.951T^{min}_{H} - 0.096^3\chi^{v}$
10	$\log L = 1.185 - 3.962\mu^2/MW - 18.252P^{max}_{\sigma-\pi} - 7.581q^{min}_{C} + 3.107V^{min}_{H} - 0.099^3\chi$
11	$\log L = 0.814 - 4.088\mu^2/MW - 15.794P^{max}_{\sigma-\pi} - 0.032HBCA + 4.183\bar{V}_H - 0.806FPSA^{(2)}$
12	$\log L = 4.447 - 0.050RPCS - 0.069RNCS + 13.962FNSA^{(2)} + 1.503RPCG + 20.263N^{max}_{C}$
13	$\log L = -33.715 - 2.449^{HA}FPSA^{(2)} + 37.940P^{max}_{\sigma-\sigma} + 0.097WNSA^{(3)}$
14	$\log L = -5.452 - 2.125\mu^2/MW + 0.920T^{min}_{H} + 1.424S^{ZX}_{ZX}$
15	$\log L = 3.032 - 0.056^2\kappa + 0.266P^{max}_{\pi-\pi} - 1.525q^{min}_{A} - 0.0062WNSA^{(1)} - 1.792RPCG$
16	$\log L = 3.086 - 1.078RNCG - 0.655FNSA^{(1)} - 0.470\Delta C_p{}^{vib}/N_A + 0.701q^{max}_{Cnet} + 0.552q^{max}_{net}$
17	$\log L = 4.584 - 0.736\Delta C^{tot}_{p}/N_A - 0.043RPCS - 0.206\nu^{h}_{TD} + 0.0029WNSA^{(2)}$
18	$\log L = 3.216 - 1.944RNCG - 0.014S_{YZ}$
	Hydrocarbons: Aromatic
19	$\log L = -1.242 - 11.658^{HA}HDCA_{\sqrt{TMSA}}{}^2 - 5.030{*}10^{-5}{*}1/6\gamma + 0.954\Delta E^{min}_{exc}(H-C) - 0.0019\Delta S_{int}/N_A + 0.472FNSA^{(2)}$
20	$\log L = 11.330 - 0.0032PPSA^{(3)} - 51.783^{HA}HDCA^{(1)}_{TMSA} - 0.365q^{min}_{net} - 8.064P_{\sigma}{}^{max}{}_{-\sigma} + 12.972N^{max}_{C} + 0.0016PNSA^{(2)}$
21	$\log L = -3.443 + 0.0041PNSA^{(2)} - 0.245^{HD}CPSA^{(2)} + 3.778\varphi^{max}_{bond}$
22	$\log L = 5.274 - 15.713FNSA^{(3)} - 0.355J + 14.867\bar{E}_C$
23	$\log L = 3.046 - 23.940FNSA^{(3)} + 0.647RPCG + 1.362S_{YZ}/R_{YZ}$
24	$\log L = 5.608 + 9.012q^{max}_{C} + 4.009q^{min}_{A} - 1.208FNSA^{(2)}$
25	$\log L = 11.624 - 1.939^{HA}HDSA_{\sqrt{TMSA}}{}^2 + 0.035\Delta H^0_f/N_A + 0.370E^{HOMO} - 2.294q^{min}_{A} + 0.0068^{HA}PSA^{(2)}$
26	$\log L = 10.521 - 83.121^{HA}HDCA^{1}_{TMSA} - 0.147\Delta E^{LUMO}_{HOMO} + 1.163\mu^2/MW - 0.364J - 2.670q^{min}_{A}$
27	$\log L = -3.535 - 0.424E^{LUMO} - 9.036HASA_{TMSA}{}^2 + 3.250\varphi^{max}_{bond} - 40.327N^{max}_{C} - 0.959RNCG$
28	$\log L = 9.303 - 2.121FHBSA + 0.079k^{max}_{A} - 0.222J + 0.255E^{HOMO} - 0.038\Delta H^0_f/N_A$
29	$\log L = 15.060 - 0.362E^{LUMO} - 0.876HASA_{\sqrt{TMSA}}{}^2 + 0.033\Delta H^0_f/N_A + 0.460E^{HOMO} - 37.757R^{max}_{C}$
30	$\log L = 9.968 - 36.362FHDCA - 0.183E^{LUMO} - 0.448J - 3.274q^{min}_{A} + 0.056E^{HOMO}$
31	$\log L = 11.178 - 0.259\Delta E^{LUMO}_{HOMO} + 0.962^0\overline{IC} + 7.583q^{min}_{C} - 0.011S_{XY} - 0.302^{HA}HDCA^{(2)}$
32	$\log L = 4.902 - 1.563\Delta C^{tot}_{P}/N_A - 0.470\Delta E^{LUMO}_{HOMO} + 5.559FPSA^{(3)} + 2.176\Delta E^{min}_{exc}(H-C)$
33	$\log L = -4.937 - 0.417E^{tot}_{R} + 0.102E^{min}_{e-n}(H-C) + 0.011^1SIC + 0.378\mu$
34	$\log L = 16.224 + 0.048^{1-center}E^{tot}_{e-e} + 0.509\mu^{tot}_{pch} + 0.866E^{tot} - 2.670^{HD}FPSA^{(2)} - 14.613\bar{P}_C$
35	$\log L = 15.894 - 5.789^{HD}FPSA^{(2)} - 0.413E^{max}_{e-e}(H-C) + 0.102\mu + 1.037\Delta E^{min}_{exc}(H-C) + 0.033^2SIC$
36	$\log L = 4.452 - 6.311FNSA^{(2)} + 5.800P + 22.283q^{min}_{C} + 33.704\bar{E}_C$
37	$\log L = 12.289 - 0.0029DPSA^{(1)} - 0.108PNSA^{(3)} + 0.331E^{HOMO}$
38	$\log L = 3.120 - 0.040^3\chi^{v} + 0.328^1\overline{IC} + 10.915\bar{E}_C + 0.0043S_{YZ}$
39	$\log L = 15.126 - 0.141E^{LUMO} - 30.959^{HD}FCPSA^{(2)} - 0.123E^{max}_{e-n}(H-C) - 0.093J$
40	$\log L = 17.434 + 0.221\Delta H^0_f/N_A + 0.523^3\chi - 6.517\mu^2/MW + 1.304N^{occ.el.lev}/N_A + 0.923^0\overline{IC}$
	Alcohols
41	$\log L = 2.017 + 1.262\mu_{hyb}{}^{tot} + 1.705I_B + 1.970P^2_f + 20.610q^{min}_{C} + 1.888q^{max}_{Cnet}$
42	$\log L = 3.146 + 1.096\mu^{tot}_{hyb} + 1.756P^2_f + 19.081q^{min}_{C} + 1.494q^{max}_{Cnet} - 0.0029\overline{\Delta H^{vib}_f}/N_A$
43	$\log L = 1.326 + 4.172P_f + 0.432\mu^{tot}_{hyb} + 2.350I_B + 1.226FNSA^{(1)} + 0.246^2\overline{IC}$
44	$\log L = 2.931 - 0.148PNSA^{(3)} - 0.313^2\overline{CIC} + 1.185\mu^2/MW - 0.022{*}1/2\beta$
45	$\log L = 8.145 + 0.963\mu^{tot}_{hyb} + 6.740q^{max}_{net} - 0.125\Delta S_{trans} + 21.103q^{min}_{C} - 36.359\bar{N}_C$
46	$\log L = 2.379 - 118.172FNSA^{(3)} + 28.234\bar{E}_C + 0.0073\Delta C^{int}_{p}$
47	$\log L = 7.707 + 4.786q^{max}_{net} + 0.841\mu^{tot}_{hyb} + 48.146I_B/N_A + 2.132\varphi^{max}_{antibond} + 0.301\nu^{h}_{TD}$
48	$\log L = 15.196 - 11.901n^{min}_{A} + 40.323I_C/N_A - 7.265 \times 10^{-4}\Delta H_{vib}/N_A$
49	$\log L = 9.146 + 1.129\mu^{tot}_{hyb} - 0.161\Delta S_{transl} + 3.156q^{max}_{net}$
50	$\log L = 3.559 + 15.153q^{max}_{A} - 0.424^2\overline{CIC} - 0.0012\Delta H^{vib}_f/N_A$
51	$\log L = 4.324 - 0.078PNSA^{(3)} + 20.362\bar{E}_C - 0.232J$
	Organic Bases
52	$\log L = 1.826 + 0.045RPCS - 0.107E^{LUMO} + 2.561\overline{I_A/N_A}$
53	$\log L = 1.894 + 0.037RPCS + 28.428I_C/N_A + 1.350^0\overline{CIC}$
54	$\log L = 2.395 + 81.001P^{max}_{\sigma-\pi} + 0.189\mu + 62.384I_C/N_A$
55	$\log L = 21.751 - 22.136FNSA^{(3)} + 25.473\bar{E}_C + 0.106^3\chi - 0.178T_C{}^{min}$
56	$\log L = 9.420 - 0.21\Delta E^{LUMO}_{HOMO} - 0.607\Delta E^{tot}_{exc} + 20.037P^{max}_{\sigma-\pi} - 0.178\Phi - 2.177\mu^2/MW$
57	$\log L = 8.722 - 35.133^{HA}HDCA_{\sqrt{TMSA}}{}^2 - 0.377T^{pairs}_{E} + 0.041^{HA}CPSA^{(2)} + 10.620E^{min}_{C}$
58	$\log L = 13.166 + 9.436P - 0.138^2\chi + 0.244E^{HOMO-1} - 3.295FPSA^{(1)} + 3.223q^{min}_{C}$
59	$\log L = 2.326 + 10.082P + 0.066E^{min}_{e-e}(C) - 5.265FNSA^{(2)} - 0.040\Phi + 2.109^{HA}FPSA^{(2)}$
60	$\log L = 12.009 + 1.533n^{max}_{A} - 0.622E^{LUMO} + 4.548\bar{V}_H - 84.108N^{max}_{C}$

Table 2 (Continued)

no.	QSPR model
	Organic Acids
61	$\log L = 10.367 + 0.066\mu - 6.462q_A^{\min} - 0.015\alpha + 6.082FNSA^{(2)} - 0.225\Delta E_{HOMO}^{LUMO}$
62	$\log L = -1.723 + 1.211n_A^{\max} - 1.670RNCG + 0.102^{HA}CPSA^{(2)} + 2.550\varphi_{bond\_}^{\max}$
63	$\log L = 1.057 - 0.017PNSA^{(1)} - 1.345q_{net}^{\min} + 3.429n_A^{\max} + 0.307\mu + 80.455\bar{N}_C$
64	$\log L = 5.405 - 1.127WNSA^{(3)} - 0.043^2CIC - 2.414RNCG - 4.945^{HD}FPSA^{(2)}$
65	$\log L = 4.759 + 1.270J - 4.050^{HD}FPSA^{(2)} - 0.031\Delta S_{tot} + 0.125\Delta E_{HOMO}^{LUMO}$
66	$\log L = 9.813 + 55.444P_{\sigma-\sigma}^{\max} + 1.905n_A^{\max} + 0.328E^{HOMO} - 1.156^{HA}HDSA_{\sqrt{TMSA}}^2$
67	$\log L = 29.474 - 0.971\varphi_{antibond}^{\max} - 0.196PPSA^{(3)} - 18.017P_H^{\min}$
68	$\log L = 8.568 + 1.574N^{HA} + 1.269\mu_{hyb}^{tot} - 0.015W - 1.554FHBSA$
	Dipolar Aprotic Species
69	$\log L = 3.890 - 1.635FPSA^{(1)} - 0.073^2\chi^v + 7.645FPSA^{(3)}$
70	$\log L = 2.931 + 0.0057PNSA^{(1)} - 0.0061PPSA^{(3)} + 11.787^{HA}FCPSA^{(2)} - 0.0012\Delta H_f^{vib}/N_A + 0.224q_{Cnet}^{\max}$
71	$\log L = 4.634 - 0.153E^{LUMO} + 0.046RPCS - 0.038\Delta S_{rot}$
72	$\log L = 6.731 - 0.142E^{LUMO} + 0.039RPCS - 0.0069PNSA^{(3)} - 0.069\Delta S_{trans}$
73	$\log L = 3.408 + 50.926I_B/N_A + 1.962^1\overline{BIC} + 0.133\mu_{p-ch}^{tot}$
74	$\log L = 2.925 + 2.122\Delta C_p^{rot}/N_A - 6.864FNSA^{(2)} + 23.351q_A^{\max}$
75	$\log L = 35.846 - 5.550RNCG + 0.033PPSA^{(3)} - 0.189\Delta E_{HOMO}^{LUMO} + 0.481\Delta C_p/N_A - 0.840E_{n-n}^{\min}(H-C)$
76	$\log L = 5.157 + 7.258q_C^{\min} - 0.080\Delta S_{trans} + 0.287P_{\pi-\pi}^{\max} + 0.0060PNSA^{(1)} - 2.807q_A^{\min}$
77	$\log L = 3.055 + 34.448\bar{E}_C - 6.566FNSA^{(2)} - 0.0021\Delta H_f^{vib}/N_A + 0.042\mu_{p-ch}^{tot} + 0.164\nu_{TD}^h$
78	$\log L = 4.799 + 0.185\Delta S_{rot}/N_A - 2.530n_A^{\min} + 0.018RPCS - 0.0018\Delta H_f^{vib}/N_A + 14.982N_C^{\max}$
79	$\log L = 3.912 - 0.137E^{LUMO} + 0.088RPCS - 0.014\Phi$
80	$\log L = 29.191 - 0.067^2\kappa + 0.181\mu_{p-ch}^{tot} + 0.225P_{\pi-\pi}^{\max} - 26.836P_{\sigma-\sigma}^{\max}$

solvents. This can be explained by the smaller quantity of uniform experimental data available for constant solute series and, as a consequence, the highly variable origin of the data generated. For instance, solubility data provided by water−solvent partition measurements as logarithms of the Ostwald solubility, by liquid−liquid chromatography in the form of infinite dilution activity coefficients, and by analytical chemistry methods in the form of molarities differ significantly and are not completely comparable. In the case of solvents, the situation is more favorable because the scattered pattern that results from nonuniform data is often compensated by extensively measured values.

For the QSPR models derived, R^2 varies from 0.604 for toluene to 0.996 for *n*-propylamine. Twenty-one solutes (25% overall) have QSPR models with R^2 less than 0.9. However, variances or squared standard deviations vary in a more narrow range. Analysis of variances, s^2, shows that the predictive ability of the models changes from excellent (0.0006) in *n*-propylamine to admissible (0.368) in piroxicam. Only 4 of 80 models have variances exceeding 0.4 kcal/mol, the generally accepted value of experimental uncertainty.[32]

Only 15 of the 80 solutes treated were previously studied theoretically by other authors: anthracene, phenanthrene, pyrene, acenaphthene, fluoranthene *trans*-stilbene, benzil, thianthrene, thioxanthen-9-one, diphenyl sulfone, hexachlorobenzene, ferrocene, fullerene, diuron, monuron, and 2-hydroxybenzoic acid, as discussed in the Introduction. The present correlations are the first for the remaining 64 solutes.

For the further discussion the QSPR models are organized according to solvent class. Definitions and discussions of the most pertinent descriptors are given throughout the text, and all are described fully in Table 4 of the Supporting Information.

Aliphatic Hydrocarbons and Chlorocompounds. As a general trend observed in Table 1, we note that the statistical quality of the QSPR models is higher for branched and cyclic hydrocarbons than for normal alkanes. Thus, *n*-octane shows a squared correlation coefficient $R^2 = 0.884$, while its isomers 2,3,4-trimethylpentane, 2,5-dimethylhexane, and ethylcyclohexane have R^2 values equal to 0.942, 0.923, and 0.946, respectively. Analysis of the variances also supports this trend: 0.037 against 0.014, 0.020, and 0.015, respectively. The descriptors selected for eqs 1−14 do not allow a uniform way of reasoning the solvating properties of different solvents with respect to alkane solutes. The most frequently occurred descriptors are geometrical, electronic, and MO-derived ones, with the minor contribution from topological indices.

Chlorosubstituted alkanes demonstrate less encouraging correlation results; this is probably due to the above-mentioned diversity of the original solubility data and the strongly nonlinear character of the intense polar interactions within the solvent media. The description of the possible nonlinear character of polar interactions with descriptors that account for nonlinearity and application of methods (neutral networks, etc.) that also account for nonlinearities will be the subject of a future study.

Electrostatic factors are of primary importance in the solubility of chlorocompounds. Each of eqs 15−18 contains either atomic charges (net or relative) or polar surface area descriptors. Among the more important quantum chemical descriptors, we mention the maximum $\pi-\pi$ bond order, $P_{\pi-\pi}^{\max}$, present in eq 15 for dichloromethane, and the highest vibrational frequency of the transition dipole, ν_{TD}^h, which occurs in eq 17 for carbon tetrachloride. The MOPAC calculated heat capacity normalized by the number of atoms in the molecule contributes negatively to the equation for highly chlorinated species such as chloroform and carbon tetrachloride, in eqs 16 and 17. Topological features of chlorocompounds are of lesser importance in their characterization; the only topological index found is the Kier shape index of the second-order, $^2\kappa$, in eq 15.

Aromatic Hydrocarbons and Halogenated Aromatics. The correlation results for aromatic and heteroaromatic solutes range from modest for toluene ($R^2 = 0.604$; $R^2_{cv} =$

0.355) to excellent for diphenyl sulfone ($R^2 = 0.977$; $R^2_{cv} = 0.966$). It is still difficult to account for all the driving forces exerted on the solvation of aromatic compounds. This is perhaps because of significant interplay of the different effects of conjugation, hyperconjugation, induction, polar resonance, etc. Again, as in the case of aliphatic chlorocompounds, electrostatic descriptors play the major role. Partial surface areas of different types are terms in almost all the QSPR equations corresponding to this class of compounds (eqs 19−40). Hydrogen bond descriptors are also important in these equations. Descriptors of the HDCA type (hydrogen bond donor charged surface area) and HASA type (hydrogen bond acceptor surface area) are present in eqs 19, 20, and 25−31. Evidently, a solvent's hydrogen bonding ability is important in the solvation of aromatic and polynuclear aromatic hydrocarbons. Quantum chemical descriptors, coding the propensity of compounds to dispersion interaction, such as the HOMO (E^{HOMO}) and LUMO (E^{LUMO}) energies and the HOMO−LUMO energy gap (ΔE^{LUMO}_{HOMO}), appear in 10 of 22 equations. In accordance with their physical meaning, E^{HOMO} bears the positive sign in all the equations, whereas E^{LUMO} bears the negative one, with the ΔE^{LUMO}_{HOMO} being always negative.

Aromatic chlorocompounds such as chlorobenzene and hexachlorobenzene are described predominantly by electrostatic descriptors. Because of the marked tendency of chlorobenzene to participate in dipole−dipole interactions (due to its rather high dipole moment), the solvent dipole moment appears in eq 35 with a positive contribution.

Alcohols. The general trend is an increase in statistical characteristics (R^2, s^2) as the size of the alkyl radical increases. There are two exceptions: 1-hexanol ($R^2 = 0.906$) and 2-propanol ($R^2 = 0.894$). On the other hand, two other branched alcohols, 2-methyl-1-propanol and 2-methyl-2-propanol, are characterized with rather high values of R^2: 0.976 and 0.960, respectively. Methanol as a solute shows rather good results with $R^2 = 0.917$ and $s^2 = 0.076$. One can observe a steady growth of the statistical parameters through ethanol to 1-heptanol. The latter has $R^2 = 0.976$ and a rather small variance of 0.007. The QSPR results for phenol, classified with alcohols in this treatment, are rather promising ($R^2 = 0.972$ and $s^2 = 0.012$) despite its increased acidity as compared to aliphatic alcohols.

The electrostatic interactions are represented in QSPR models 41−45, 47, and 49−50 by the atomic charges and dipole moments. The electrostatic descriptors used most frequently are the hybridization component of the molecular dipole, μ^{tot}_{hyb} (6 of 11 models), and the minimum (or maximum) atomic charges on carbon atom and on a generic atom, q^{min}_C and q^{max}_A, respectively. Another electrostatic descriptor, found in eqs 41−43 for the three simplest alcohols, is the polarity parameter (P^2_f), a function of atomic charges and the squared distance between the atoms. All the electrostatic descriptors in the models corresponding to alcohols bring a positive contribution to the solubility. This observation implies that, other things being equal, the degree of transfer to the solvent phase from the gas phase is greater for high-polarity compounds.

Frontier molecular orbital indices such as the average electrophilic reactivity index for carbon atoms ($\bar{E}_C$), the average nucleophilic reactivity index for carbon atoms ($\bar{N}_C$), and the minimum atomic orbital electronic population (n^{min}_A) are also important in the description of dispersion forces and other weak solvation effects.

Contributions from cavity-forming and dispersion forces, which are also significant in alcohol solutions, are reflected by a set of geometrical and topological descriptors such as the Balaban index (eq 51), the information topological indices $^2\overline{IC}$ and $^2\overline{CIC}$ (eqs 43, 44, and 50), and the moments of inertia along axes B or C (eqs 41, 43, 47, and 48).

Organic Bases. The best correlations are obtained for simple aliphatic amines. Ethyl-, *n*-propyl-, and *n*-butylamine have $R^2 = 0.974$, 0.996, and 0.978, respectively. The variance value for *n*-propylamine is extremely low at 0.0006. Aniline, the simplest aromatic amine, also has good statistical features, $R^2 = 0.956$ and $s^2 = 0.032$. For the two *para*-nitrocompounds, 4-nitroaniline and 4-nitro-*N*,*N*-dimethylaniline, the correlation results are not so high, especially for 4-nitroaniline: $R^2 = 0.806$. The complex nature of the electronic charge distribution (driven by the pronounced resonance conjugation between the nitro and amino groups) combines in 4-nitroaniline, with a possible implication for hydrogen bond formation and proton-transfer processes.

The solvation processes of the nitrogen bases appear to be dominated by electrostatic, dispersion, and hydrogen bond forces encoded in descriptors presented in Table 2, lines 52−60. The dipole moment μ or its function, the image of the Onsager-Kirkwood solvation energy (μ^2/MW), are terms in eqs 54 and 56. Other electrostatic descriptors are the relative positive charged surface area (*RPCS*), the partial atomic charge on a carbon atom (q^{min}_C), the polar partial surface areas such as $FNSA^{(3)}$, $FPSA^{(1)}$, and $CPSA^{(2)}$, and the topographic electronic index over all atoms' pairs (T^{pairs}_E).

Dispersion forces can be related to the MO entities such as the LUMO energy (eqs 52 and 60), the HOMO-1 energy, the HOMO−LUMO energy gap, and the maximum atomic orbital electronic population (n^{max}_A). As an illustration of hydrogen bonding descriptors we refer to the *H*-acceptor dependent HDCA-2 ($^{HA}HDCA^{(2)}$), the new *H*-donors charged partial surface area ($^{HA}CPSA^{(2)}$), and the maximum $\sigma-\pi$ bond order, which can be loosely related to the basicity of the solvents under study. The bulk properties of the solvents are represented in eqs 52−56 and eqs 58−59 by the moments of inertia (along axes A and C), the Randic indices of order 2 and 3, and the Kier flexibility index.

Organic Acids. Of the eight solutes with an acidic nature, four are benzoic acid and substituted benzoic acids, one is a derivative of phenylacetic acid (diclofenac), one is a derivative of benzeneacetic acid, and two are rather acidic hydroxyaromatic compounds (haloperidol and paracetamol). The QSPR models for ibuprofen and diclofenac are of poor statistical quality (R^2_{cv} is equal to 0.549 and 0.579, respectively), probably because of the complex multifunctional structure of these solutes; another relevant factor could be the small number of experimental data points used in the modeling (24 and 23, respectively). A small data set is also the likely reason for the significant gap between R^2_{cv} and R^2 in the case of haloperidol: 0.934 vs. 0.815 with only 17 data points. The best five-parameter model was obtained for 4-hydroxybenzoic acid ($R^2 = 0.936$; $R^2_{cv} = 0.862$; $s^2 = 0.094$).

The effects of the acidic nature on the solvation of the solutes under discussion is apparently reflected in eqs 61−68 by the participation of solvent electrostatic descriptors such as the following: the dipole moment (3 of 8 models), the partial and relative atomic charges (q_A^{min} and *RNCG*), and the polar surface areas (*FNSA*(2), *PNSA*(1), *WNSA*(3), and *PPSA*(3)). Hydrogen bonding patterns, which are very important in the solvation of organic acids, are represented in eqs 62 and 64−68 by the following hydrogen bond descriptors: the new *H*-donor charged partial surface area ($^{HA}CPSA^{(2)}$), the new *H*-donor fractional partial positive surface area ($^{HD}FPSA^{(2)}$), the *H*-acceptor dependent HDSA-2, ($^{HA}HDSA^{(2)}$), the fractional *H*-bonding surface area (HBSA/TMSA), and the count of hydrogen acceptor sites (N^{HA}). Cavity-forming and dispersion forces are coded by molecular polarizability, the energy of the HOMO level and the gap between the HOMO and LUMO, and a set of topological indices such as the Balaban index (*J*), the Wiener index (*W*), and the second-order complementary information content (^2CIC).

Dipolar Aprotic Species. Twelve dipolar aprotic species are represented by the following solvents, lines 69−80 of Tables 1 and 2: 6 esters, 3 ketones, 1 nitrile, 1 nitrocompound, and 1 six-membered ring ether. All of the solutes are characterized by good to excellent statistical models. The statistically highest correlation coefficients were shown for 2-hexanone ($R^2 = 0.993$; $R^2_{cv} = 0.990$; $s^2 = 0.001$); lower but convincing results were obtained for 2-butanone ($R^2 = 0.913$; $R^2_{cv} = 0.871$; $s^2 = 0.017$). We note the low values of the variance (s^2, the squared standard deviation): just two equations display large s^2 values (0.047 and 0.034 for acetonitrile and nitromethane, respectively).

The distribution of electronic charge, expressed in terms of the partial positively or negatively charged surface area and atomic charges, plays a key role in the solvation of aprotic dipolar solutes. Ten of 12 QSPR models derived include different partial surface areas such as *FPSA*(1), *FPSA*(3), *PNSA*(1), *PPSA*(3), and *FNSA*(2). The superscripts indicate the type of atomic charges used in weighting the polar surface areas.[33] Partial and relative atomic charges (q_{Cnet}^{max}, q_A^{max}, q_A^{min}, q_C^{min}, *RPCS*, and *RNCG*, see Table 4 for keys for descriptors) are found in 8 of 12 equations. The dipole influence is accounted for by the point charge component of the molecular dipole (μ_{p-ch}^{tot}) in eqs 73, 77, and 80. As for the solvent bulk properties, which exert influence on the dipolar species phase distribution, the entropic Kier−Hall valence connectivity index of order 2 ($^2\chi^v$), the Kier−Hall flexibility index (Φ), the Kier shape index of order 2, and the first-order average bonding information content ($^1\overline{BIC}$). In eq 73, the moment of inertia along axis B is chosen as an additional cavity-formation term, reflecting the complex geometry and the high flexibility of pentyl acetate.

GENERAL CONCLUSIONS

A pool of approximately 800 descriptors was analyzed using the heuristic method to give correlation equations for 80 solutes in a range of solvents. Sixty-four of the solutes are studied by QSPR methodology for the first time. The predictive quality of several equations suffers from a significant degree of clustering in the data sets: for 40 of the solutes (lines 2, 4−6, 10, 12−25, 29, 31−32, 35, 37−38, 44, 52, 56, 60−68, 75, and 80 in Table 1) differences between the R^2 and R^2_{cv} values are higher than 0.05 in R^2 units. Most probably the data clustering is due to experimental uncertainties of the measurements and the nonuniform character of some particular data sets (e.g. many nonpolar species and few polar alcohols or acids).

The descriptive quality of the equations is in good agreement with our understanding of solute−solvent interactions, as illustrated in this paper and in previous work. The descriptors contained in the equations emphasize nonspecific interactions between solute and solvent, which are driven by the dipole−dipole interactions, hydrogen bond donor/acceptor functionality, and bulk-related properties of solute and solvent molecules. Thus, the descriptors reflect the electronic charge distribution, surface area, and various other structural properties of the compounds.

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 for series of solutes (Table 3 (3-1−3-80)) and explanation of the descriptors used in Table 2 (Table 4). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) 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.

(2) 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.

(3) 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.

(4) 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.

(5) Katritzky, A. R.; Oliferenko, A.; Oliferenko, P.; Petrukhin, R.; Tatham, D.; Maran, U.; Lomaka, A.; Acree, W. A General Treatment of Solubility. 1. The QSPR Correlation of Solvation Free Energies of Single Solutes in Series of Solvents. **2003**, *43*, 1794−1805.

(6) Abboud, J. L.; Taft, R. W.; Regarding a Generalized Scale of Solvent Polarities. *J. Am. Chem. Soc.* **1977**, *99*, 8325−8327.

(7) 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.

(8) Abraham, M. H. Physicochemical and Biological Processes. *Chem. Soc. Rev.* **1993**, 73−83.

(9) Acree, W. E.; Abraham, M. H. Solubility Prediction for Crystalline Nonelectrolyte Solutes Dissolved in Organic Solvents Based Upon the Abraham General Solvation Model. *Can. J. Chem.* **2001**, *79*, 1466−1476.

(10) Abraham, M. H.; Green, C. E.; Acree, W. E., Jr., Hernandez, C. E.; Roy, L. E. Descriptors for Solutes from the Solubility of Solids: *trans*-Stilbene as an Example. *J. Chem. Soc., Perkin Trans. 2* **1998**, 2677−2681.

(11) Abraham, M. H.; Benjelloun-Dakhama, N.; Gola, J. M. R.; Acree, W. E., Jr., Cain, W. S.; Cometto-Muniz, J. E. Solvation Descriptors for Ferrocene, and the Estimation of Some Physicochemical and Biochemical Properties. *New J. Chem.* **2000**, *24*, 825−829.

(12) Abraham, M. H.; Green, C. E.; Acree, W. E. Correlation and Prediction of the Solubility of Buckminsterfullerene in Organic Solvents; Estimation of Some Physicochemical Properties. *J. Chem. Soc., Perkin Trans. 2* **2000**, 281−286.

(13) Green, C. E.; Abraham, M. H.; Acree, W. E., Jr.; De Fina, K. M.; Sharp, T. L. Solvation Descriptors for Pesticides from the Solubility of Solids: Diuron as an Example. *Pest Manage. Sci.* **2000**, *56*, 1043−1053.

(14) 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.

(15) Roy, L. E.; Hernandez, C. E.; Acree, W. E. Solubility of Anthracene in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Polycyclic Aromat. Compd*. **1999**, 13, 105−116.

(16) Powell, J. R.; Voisinet, D.; Salazar, A.; Acree, W. E. Solubility of Pyrene in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Phys. Chem. Liq.* **1994**, *28*, 269−276.

(17) De Fina, K. M.; Sharp, T. L.; Acree, W. E., Jr. Solubility of Acenaphthene in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Can. J. Chem.* **1999**, *77*, 1537−1541.

(18) Roy, L. E.; Hernandez, C. E.; Acree, W. E., Jr. Thermodynamics of Mobile Order Theory. Part 3. Comparison of Experimental and Predicted Solubilities for Fluoranthene and Pyrene. *Polycyclic Aromat. Compd*. **1999**, *13*, 205−219.

(19) Fletcher, K. A.; McHale, M. E. R.; Coym, K. S.; Acree, W. E. Solubility of *trans*-Stilbene in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile order theory. *Can. J. Chem.* **1997**, *75*, 258−261.

(20) Roy, L. E.; Hernandez, C. E.; De Fina, K. M.; Acree, W. E., Jr. Thermodynamics of Mobile Order Theory. Part 4. Comparison of Experimental and Predicted Solubilities for *trans*-Stilbene. *Phys. Chem. Liq*. **2000**, *38*, 333−343.

(21) Fletcher, K. A.; Pandey, S.; McHale, E. R.; Acree, W. E. Solubility of Benzil in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Phys. Chem. Liq.* **1996**, *33*, 181−190.

(22) Fletcher, K. A.; McHale, M. E. R.: Powell, J. R.; Coym, K. S.; Acree, W. E. Solubility of Thianthrene in Organic Nonelectrolyte Solvents: Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Phys. Chem. Liq.* **1997**, *34*, 41−49.

(23) Fletcher, K. A.; Coym, K. S.; Roy, L. E.; Hernandez, C. E.; Mchale, M. E. R.; Acree, W. E. Solubility of Thioxanthen-9-one in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Phys. Chem. Liq*. **1998**, *35*, 243−252.

(24) Fletcher, K. A.; Hernandez, C. E.; Roy, L. E.; Coym, K. S.; Acree, W. E., Jr. Solubility of Diphenyl Sulfone in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon the General Solvation Model. *Can. J. Chem*. **1999**, *77*, 1214−1217.

(25) De Fina, K. M.; Ezell, C.; Acree, W. E. Solubility of Ferrocene in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Phys. Chem. Liq*. **2001**, *39*, 699−710.

(26) Huyskens, F.; Morissen, H.; Huyskens, P. Solubilities of *p*-Nitroanilines in Various Classes of Solvents. Specific Solute−Solvent Interactions. *J. Mol. Struct*. **1998**, *441*, 17−25.

(27) De Fina, K. M.; Sharp, T. L.; Spurgin, M. A.; Chuca, I.; Acree, W. E.; Green, C. E.; Abraham, M. H. Solubility of the Pesticide Diuron in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Can. J. Chem.* **2000**, *78*, 184−190.

(28) De Fina, K. M.; Sharp, T. L.; Chuca, I.; Spurgin, M. A.; Acree, W. E. Jr.; Green, C. E.; Abraham, M. H. Solubility of the Pesticide Monuron in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based upon Mobile Order Theory. *Phys. Chem. Liq.* **2002**, *40*, 255−268.

(29) Sivaraman, N.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Natarajan, R. QSPR Modeling for Solubility of Fullerene (C_{60}) in Organic Solvents. *J. Chem. Inf. Comput. Sci*, **2001**, *41*, 1067−1074.

(30) Danauskas, S. M.; Jurs, P. C. Prediction of C_{60} Solubilities from Solvent Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 419−424.

(31) Shang, Z.-C.; Zou, J.-W.; Huang, M.-L.; Yu, Q.-S. QSPR Studies on Solubilities of Some Given Solutes in Pure Solvents Using Frontier Orbital Energies and Theoretical Descriptors Derived from Electrostatic Potentials on Molecular Surface. *Acta Chim. Sin.* **2002**, *60*, 647−652.

(32) 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.

(33) 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.

CI034122X