1598 *J. Phys. Chem. A* **1999,** *103,* 1598−1610

Modeling with Semiempirical Molecular Connectivity Terms

Lionello Pogliani[†]

Dipartimento di Chimica, Universita' della Calabria, 87030 Rende (CS), Italy

Received: August 4, 1998; In Final Form: October 28, 1998

A quest for an optimal missing descriptor is started here, a descriptor that can include information on van der Waals and hydrogen bond interactions. This quest is centered on the construction of special molecular connectivity terms that make use of an empirical parameter, the dielectric constant, and three different "ad hoc" parameters indirectly related to the dielectric constant that should render molecular connectivity terms able to describe also noncovalent interactions. The molar mass *M* was also used, as an empirical parameter, to derive molecular connectivity terms strongly related to the mass of the compounds. Derived semiempirical molecular connectivity terms were used to model 11 different properties of a wide heterogeneous range of solvents, for which χ indices alone and their linear combinations are insufficient descriptors: boiling points, melting points, refractive indices, density, viscosities, flashpoint values, elutropic values, UV cutoff values, dipolar moments, magnetic susceptibility, and the same dielectric constant. Achieved modelings are quite fair, and with the exclusion of the refractive indices all of them needed semiempirical terms, mainly based on supramolecular connectivity indices of different forms. Further, a common term, within the "double sieve" approximation for the melting points, has been detected for the boiling and melting temperatures. Two types of valence molecular connectivity indices have been used all along the modeling: χ^v indices based on δ^v valence and χ^z indices based on δ^z valence numbers. This last kind of χ^z index plays an important role in *X* terms for the flashpoints and melting points only.

Introduction

Recently, supramolecular connectivity indices[1−3] and molecular connectivity terms[4−7] have been successfully introduced to model different properties of amino acids, purines and pyrimidines, alkanes, and inorganic salts. These descriptors are based on the original molecular connectivity indices developed more than 20 years ago[8−13] and further improved, studied, and used all along these years by different authors.[14−42] Normally, topological studies on molecular and bulk properties are mainly based on the prediction of properties (such as boiling points, the central property of QSPR studies) of rather homogeneous sets of compounds, where either the nearly complete disruption or the underestimation of intermolecular forces can be assumed. This explains the difficulty encountered to model, e.g., the melting points of simple alkanes, which seems to lie beyond the descriptive capability of molecular connectivity χ indices and other kinds of topological indices, constituting, thus, a kind of "hard" property. Thus, even if molecular connectivity indices generally provide the most accurate modeling for several properties of homogeneous classes of compounds, they nevertheless seemed to fail to describe some of their properties, and especially the properties of highly heterogeneous sets of molecules. The introduction of supramolecular $a\chi$ and molecular connectivity $X = f(\chi)$ terms aimed, and in some cases succeeded, to reduce this divergence between "hard" and "soft" properties. The charge to pay has clearly been to give up the kind of exceptional description attained for some "soft" properties, even if, in many cases a rather satisfactory description could be achieved.[1−7]

Noncovalent interactions are increasingly being used in the molecular self-assembly of well-defined supramolecular structures and materials. Such interactions are important in material science where weak reversible interactions are important in determining material structure and activity, and their description represents a high challenge for topological indices, and especially for molecular connectivity indices, which are dependent on the overall ramification or pseudoramification degree of molecules, where pseudoramification is understood to be the ramification in chemical pseudographs.[6] In this work, different physicochemical properties of a highly heterogeneous class of organic solvents (saturated, unsaturated, unsubstituted, slightly and highly substituted, nonpolar, and slightly polar and highly polar compounds) will be described by the aid of different molecular connectivity terms, which include one or more empirical parameters. None of these properties can satisfactorily be modeled by "pure" χ indices or *X* terms, and only the introduction into the X terms of an extra empirical parameter enhances their modeling, which seems mainly or to some extent determined by noncovalent interactions. Two different sets of valence connectivity indices will be also be checked: the χ^v indices based on the well-known δ^v cardinal number[13] and the χ^z indices based on the δ^z cardinal number, recently introduced to model some properties of amino acids and inorganic salts.[3,4] Modeled properties are: boiling points, T_b, melting points, T_m, refractive index, RI, density, *d*, flashpoint, FP, viscosity, η (or VIS), elutropic value, EV, UV cutoff value, UV, dipole moment, μ (or DM), susceptibility, χ (or SUSC), and finally, the same dielectric constant ϵ (or DM). The main goals of this paper are quite general: (i) to find one or more empirical descriptors that enhance the descriptive power of molecular connectivity indices or terms and that can give us some hints about the general structure of the missing noncovalent descriptor, (ii) to elucidate that the process of construction of semiempirical molecular connectivity terms provides consistent and useful modeling of

---

[†] Fax: 39-984-492044. E-mail: lionp@pobox.unical.it.

10.1021/jp983268+ CCC: $18.00

as many properties as possible of a wide and heterogeneous set of compounds because of the introduction of semiempirical terms, and (iii) to check how far one can take the intermixing of molecular connectivity indices and empirical parameters for a better modeling.

Method

Molecular X connectivity terms[5−7] are a kind of higher-level molecular connectivity indices that can be derived by a trial-and-error composition procedure based on a medium-sized $\{\chi\}$ set of eight optimal[1−7] molecular and valence (denoted by the lowercase v) molecular connectivity indices. The supramolecular connectivity index[1−4] can be considered the simplest form of a molecular connectivity term; it is in fact a molecular connectivity index multiplied by an association constant a. In this work, to reduce as much as possible the combinatorial problem, and the dimension of the trial-and-error procedure to derive molecular connectivity terms, two different sets of the eight most used optimal molecular connectivity indices will be used for combinatorial and trial-and-error operations

$$\{\chi, \chi^{v}\} = \{D, D^{v}, {}^{0}\chi, {}^{0}\chi^{v}, {}^{1}\chi, {}^{1}\chi^{v}, \chi_{t}, \chi_{t}^{v}\} \qquad (1)$$

$$\{\chi, \chi^{z}\} = \{D, D^{z}, {}^{0}\chi, {}^{0}\chi^{z}, {}^{1}\chi, {}^{1}\chi^{z}, \chi_{t}, \chi_{t}^{z}\} \qquad (2)$$

These indices are based on the degree δ_i (for χ) and δ_i^{v} (for χ^{v})[13] of heteroatom i represented by a point in hydrogen-suppressed graphs or pseudographs,[6] respectively. A satisfactory algorithm to compute δ_i^{v} for second and higher row atoms is[13] $\delta_i^{v} = Z^{v}/(Z - Z^{v} - 1)$, where Z^{v} is the number of valence electrons and Z is the atomic number of the heteroatoms in hydrogen-suppressed molecules (for S the value 6/9 was used; see ref 13 for many of these values). For O, N, and halogens (F, Cl, and Br) heteroatoms χ^{z} type indices[3,4] based on degree $\delta_i^{z} = Z^{v}/n$, have been introduced, where n is the principal quantum number. Thus, for N, NH, NH_2, O, OH, F, S, Cl, and Br, we have 5/2, 4/2, 3/2, 6/2, 5/2, 7/2, 7/3, 6/3, and 7/4, respectively. The advantage of the first definition is that it fits the original concept of the vertex degree in pseudographs for second row atoms.

Molecular connectivity indices of the given sets can be computed as follows: the sum-delta D index is given[28] by

$$D = \Sigma\delta_i \qquad (3)$$

The zeroth- and first-order indices are defined as[13] follows

$${}^{0}\chi = \Sigma(\delta_i)^{-0.5} \qquad (4)$$

$${}^{1}\chi = \Sigma(\delta_i\delta_j)^{-0.5} \qquad (5)$$

while the total structure molecular connectivity index, χ_t, over the N non-hydrogen atoms of the molecule[18] is

$$\chi_t = (\delta_1\delta_2...\delta_N)^{-0.5} \qquad (6)$$

Sums in eqs 3−5 run over the total number of vertexes (eqs 3 and 4) and edges (eq 5) of the hydrogen-suppressed molecular graph, respectively. Replacing δ in eqs 3−6 with valence δ^{v} and with δ^{z}, the corresponding valence molecular connectivity χ^{v} and χ^{z} indices are obtained, respectively.

The linear estimation problem of a physicochemical property P with χ indices or $X = f(\chi)$ terms is the estimation of the experimental property P_{exp} in terms of a linear function of χ or X, which in matrix notation is given by the following dot product

$$P = |\mathbf{C}\cdot\mathbf{X}| \qquad (7)$$

where $X = \chi$ depicts a special case and P is the calculated property of a compound. Row vector **C** is the vector of coefficients c_k, determined by the linear least-squares procedure, and column vector X is the vector of the best molecular connectivity X terms, derived by the aid of a combinatorial procedure, based, normally, on the complete combinatorial technique, i.e., over the total combinatorial space of the eight connectivity indices. The descriptor corresponding to the constant c_0 term is the unitary index $X^0 \equiv 1$. Bars in eq 7 stand for absolute value to get rid of negative P values with no physical meaning, simultaneously enhancing the description of the property. For properties with negative values, bars in eq 7 are omitted. Molecular connectivity terms are found with a trial-and-error procedure, which stops as soon as a higher-level descriptor with satisfactory values for the following statistics is found: $\langle u\rangle$, $Q = r/s$, $F = fr^2/[(1 - r^2)\upsilon]$, where, r = correlation coefficient, s = standard deviation of the estimate, f = degrees of freedom, υ = number of χ or X indices. For every term of the best LCXCT (linear combination of X molecular connectivity terms) equation the utility, $u_k = |c_k/s_k|$ of every index, as well as the average utility $\langle u\rangle = \sum u_k/m$ of the indices of the found linear combination will be estimated. The same statistics are also used to derive the best linear combination of χ indices, which are in many cases used to construct the X terms. An interesting feature of these terms is that, in many occasions, they show dominant character, a fact that helps to restrict the combinatorial search for a multilinear description.[3,5,6] In fact, it is not rare the case, when X terms are used in linear combinations with $\{\chi\}$ indices or $\{a\chi\}$ supraindices, that the complete combinatorial search changes over to a forward selection search, where the search is restricted to the next best index only, simplifying, thus, the combinatorial problem. Molecular connectivity terms can also be orthogonalized, generating orthogonal molecular connectivity terms; this fact (i) short-circuits the collinearity problem due to the mutual interrelation among the X terms, (ii) improves $\langle u\rangle$ and u_k, (iii) generates coefficients that are stable upon introduction of a new orthogonal index, and (iv) detects dominant descriptors whenever normal indices are poor descriptors.[20−22,35] Clearly, the trial-and-error procedure could give rise to an infinite form of terms, even with a medium sized set of $\{\chi\}$ indices. To avoid such a pitfall, the construction of molecular connectivity terms will be centered around some well-defined forms, starting with the well-known general form, which has already been very successful in modeling different properties of biochemical compounds,[5−7]

$$X = \chi_1/(\chi_2 + b\chi_3) \qquad (8)$$

where b (≥ 0 or <0) is an optimization constant and where two indices can even be equal and one equal to $1 \equiv \chi^0$. Parameter b should not be confused with the association parameter a, used to construct supramolecular connectivity indices, which is normally a parameter inferred or deduced from experimental evidence of association phenomena in solution.[1,2,33] Sometimes one or more χ_i indices can have an exponent. The building of such terms is rather easy, just start with two indices, cyclically optimize each index over the set of ≤ 8 indices (sign "<" if only best χ indices are chosen), introduce then the third index, cycling again, then introduce the b parameter and exponents and optimize again. The procedure, even if a little bit tedious, is rather straightforward, and the chances of success are rather high. This procedure is reminiscent of the shimming procedure

TABLE 1: Physicochemical Properties of Organic Solvents[a]

solvent	M	T_m	T_b	RI	d	ϵ	FP	η	EV	UVco
acetone	58.1	−94	56	1.3590	0.791	20.7	−17	0.32	0.43	330
acetonitrile	41.05	−48	82	1.3440	0.786	37.5	5	0.37	0.50	190
benzene	78.1	5	80	1.5010	0.874	2.3	−11	0.65	0.27	280
benzonitrile	103.1	−13	188	1.528	1.010	25.2	71	1.24[1]		
1-butanol	74.1	−90	117.7	1.3990	0.810	17.1	35	2.948		215
2-butanone	72.1	−87	80	1.3790	0.805	18.5	−3	0.40	0.39	330
butyl acetate	116.2	−78	125	1.3940	0.882	5.01	22	0.732		254
CS_2	76.1	−112	46	1.6270	1.266	2.6	−33	0.37		380
CCl_4	153.8	−23	77	1.4595	1.594	2.2	-	0.97	0.14	263
Cl−benzene	112.6	−45	132	1.5240	1.107	5.62	23	0.80		287
1Cl−butane	92.6	−123	77.5	1.4024	0.886	7.39	−6	0.35		225
$CHCl_3$	119.4	−63	61	1.4460	1.492	4.8	-	0.57	0.31	245
cyclohexane	84.2	6.5	80.9	1.426	0.779	2.0	−18	1.00	0.03	200
cyclopentane	70.1	−94	50	1.4000	0.751	2.0	−37	0.47		200
1,2-dCl−benzene	147.0	−16.5	179.5	1.5510	1.306	9.9	65	1.324		295
1,2-dCl-ethane	98.95	−35	83	1.4438	1.256	10.37	15	0.79		225
dCl-methane	84.9	−97	39.9	1.4240	1.325	7.5		0.44	0.32	235
N,*N*-dM-acetamd	87.1	−20	165.2	1.4380	0.937	37.8	70			268
N,*N*-dM-formamd	73.1	−61	153	1.431	0.944	36.7	57	0.92		268
1,4-dioxane	88.1	11.8	101	1.4220	1.034	2.2	12	1.54		215
ether	74.1	−116	34.6	1.3530	0.708	4.3	−40	0.23	0.29	215
ethyl acetate	88.1	−84	77	1.3720	0.902	6.0	−3	0.45	0.45	260
ethyl alcohol	46.1	−130	78	1.3600	0.785	24.5	8	1.20		210
heptane	100.2	−91	98	1.3870	0.684	1.92	−1		0.00	200
hexane	86.2	−95	69	1.3750	0.659	1.89	−23	0.33	0.00	200
2-methoxyethanol	76.1	−85	124.5	1.4020	0.965	16.0	46	1.72		220
methyl alcohol	32.0	−98	64.6	1.3290	0.791	32.7	11	0.60	0.73	205
2-methylbutane	72.15	-	30	1.3540	0.620	1.843	−56			
4-M-2-pentanone	100.2	−80	117.5	1.3960	0.800	13.1	13			334
2-M-1-Propanol	74.1	−108	108	1.3960	0.803	17.7	37			
2-M-2-Propanol	74.1	25	83	1.3870	0.786	10.9	4			
DMSO	78.1	18.4	189	1.4790	1.101	46.7	95	2.24		268
nitromethane	61.0	−29	100.9	1.3820	1.127	35.9	35	0.67		380
1-octanol	130.2	−15	196	1.4290	0.827	10.34	81			
pentane	72.15	−130	35.5	1.3580	0.626	1.844	−49	0.23	0.00	200
3-pentanone	86.1	−40	102	1.3920	0.853	17.0	6			
1-propanol	60.1	−127	97	1.3840	0.804	20.1	15	2.256		210
2-propanol	60.1	−89.5	82.4	1.3770	0.785	18.3	22	2.30	0.63	210
pyridine	79.1	−42	115	1.5100	0.978	12.4	20	0.94	0.55	305
ttCl-ethylene	165.8	−22	121	1.5056	1.623	2.3		0.90		
tt-hydrofuran	72.1	−108	67	1.4070	0.886	7.6	−17	0.55	0.35	215
toluene	92.1	−93	111	1.4960	0.867	2.4	4	0.59	0.22	285
1,1,2tCltFEthane	187.4	−35	47.5	1.3578	1.575	2.41		0.69	0.02	230
2,2,4-tM-pentane	114.2	−107	98.5	1.3910	0.692	1.94	−7	0.50	0.01	215
o-xylene	106.2	−24	144	1.5050	0.870	2.568	32	0.810		
p-xylene	106.2	12.5	138	1.4950	0.866	2.374	27	0.648		
acetic acid	60.05	16.7	117.9	1.3719	1.049	6.1				
decaline	138.2	−43	191.7	1.4758	0.879	2.20				
dBr-methane	173.8	−52.6	97.0	2.4970	1.542	7.5				
1,2-dCl-E-en(Z)	96.9	−80.0	60.6	1.4490	1.284	9.2				
1,2-dCl-E-en(E)	96.9	−49.8	47.7	1.4462	1.255	2.1				
1,1-dCl-E-en	96.9	−122.6	31.6	1.4247	1.213	4.6				
dMethoxymethane	76.1	−105.2	42.3	1.3563	0.866	2.6				
dMethyl ether	46.1	−139	−24			5.02				
E-enCarbonate	88.1	36.4	238	1.4250	1.321	89.6				
formamide	45.0	2.6	210.5	1.4475	1.133	109				
methyl chloride	50.5	−97.7	−24.1	1.3389	0.916	12.6				
morpholine	87.1	−3.1	128.9	1.4573	1.005	7.4				
quinoline	129.2	−14.9	237.1	1.6293	1.098	9.0				
SO_2	64.1	−72.7	−10.0		1.434	17.6				
2,2-ttCl-Ethane	167.8	−43.8	146.2	1.4868	1.578	8.2				
ttM-Urea	116.2	−1.2	176.5	1.4493	0.969	23.1				
tCl-E-en	131.4	−73	87.2	1.4800	1.476	3.4				

[a] M = molar mass (g·mol^{-1}), T_m = melting points (°C), T_b = boiling points (°C), RI = refractive index (20 °C), d = density (at 20 ± 5 °C relative to water at 4 °C), ϵ = dielectric constant, FP = flash point (°C), η = viscosity (cpoise, 20 °C; [1] at 25 °C), Elutropic value (silica), UVco = UV cutoff (nm, wavelength at which absorbance is 1 Å for a good LC grade solvent). Properties are taken from refs 42−44. Abbreviations: amd = amide, d = di, E-en = ethylen, M = methyl, t = tri, tt = tetra. Both T_b and T_m have been modeled in Kelvins.

for magnetic field optimization in NMR, where a series of small field gradients (about 12 on a Bruker 600 MHz) are cyclically optimized, mostly by hand, to obtain a highly homogeneous magnetic field.

Results and Discussion

The different physicochemical properties of the given solvents are collected in Tables 1 and 2. They have been obtained from

TABLE 2: Dipole Moments, μ, in Debye Units (1 D = 10^{-18} esu cm = 3.3356 × 10^{-3} C m) and Magnetic Suceptibility in $-10^6\chi$ of 44 Organic Solvents, at T = 20 °C, unless Otherwise Written

Solvent	acetic acid	acetone	acetonitrile	benzene
μ	1.2	2.88	3.92	0
$-10^6\chi$	0.551 (32°)	0.460	0.534	0.699 (22 °C)
Solvent	2-M-2-propanol	CS_2	CCl_4	$CHCl_3$
μ	1.66	0	0	1.01
$-10^6\chi$	0.534	0.532	0.691	0.740
Solvent	cyclohexane	cyclopentane	decaline	dBr-methane
μ	0			1.43
$-10^6\chi$	0.627	0.629	0.681	0.935
Solvent	1,2-dCl-benzene	ether	1,2-dCl-ethane	1,2-dCl-E-en(Z)
μ	2.50	1.15	1.75	1.90
$-10^6\chi$	0.748			0.679 (15°)
Solvent	1,2-dCl-E-en(E)	1,1-dCl-E-en	dCl-methane	dMethoxymethane
μ	0	1.34	1.60	
$-10^6\chi$	0.638 (15°)	0.635 (15°)	0.733	0.611
Solvent	*N*,*N*-dM-acetamd	*N*,*N*-dM-formamd	dMSulfoxide	1,4-dioxane
μ	3.8	3.86	3.96	0.45
$-10^6\chi$				0.606 (32°)
Solvent	ethyl alcohol	ethyl acetate	E-enCarbonate	formamide
μ	1.69	1.8	4.91	3.73
$-10^6\chi$	0.575	0.554		0.551
Solvent	methyl alcohol	methyl chloride	morpholine	nitromethane
μ	1.70	1.87		3.46
$-10^6\chi$	0.530		0.631	0.391 (25°)
Solvent	pyridine	quinoline	SO_2	2,2-ttCl-ethane
μ	2.2	2.2	1.6	1.3
$-10^6\chi$	0.611	0.729		0.856
Solvent	ttCl-ethylene	tt-hydrofuran	ttM-urea	toluene/tCl-E-en
μ		1.75	3.47	0.36/−
$-10^6\chi$	0.802 (15°)		0.634 (15°)	0.618 / 0.734

the available literature.[43−45] Values of molecular connectivity indices, χ, χ^v, and χ^z for these compounds are collected in Tables 3 and 4. It should here be noticed that both T_m and T_b have been modeled in Kelvin degrees (T/K = T/°C + 273.15). In Table 4 are collected only χ^z values because the normal molecular connectivity χ values are unique, as can be seen from eqs 3−5. Table 1 shows also the values of the dielectric constants of the different solvents, a central property in this study. This property, either directly or indirectly, together with 'ad hoc′ ϵ-related parameters, which will describe hydrogen bonds in alcohols and acids, will here be used to overcome the inherent limitation of the molecular connectivity indices, which do not encode van der Waals and/or hydrogen bond interactions. Molar masses, M, will also be used to improve the description whenever required. The introduction of such 'ad hoc', M, ϵ, and ϵ-related parameters will also allow us to analyze some general characteristics (bulk or noncovalent) of the best descriptors. The dielectric constant has been chosen to improve the modeling as (i) it is related to the noncovalent character of a compound, (ii) a wide wealth of values for this property are given (see Table 2), and (iii) one can follow what is normally done in molecular dynamics simulations, where the solvent is normally mimicked by using its dielectric constant. From ϵ some special weak intermolecular association parameters have been derived, the first one being $a_w \approx \epsilon/15$ parameter (truncated at the first figure), while the other two (the rather similar, a_{OH} and a_ϵ) will be defined all along the paper. For $\epsilon/15 < 1$, $a_w = 1$ will be assumed (see Table 4). Number 15 has been chosen as it represents the molar mass of a CH_3 radical, the most simple radical in organic chemistry. Hydrogen bonds in alcohols and acetic acid contribute $a_w = 2$, whatever the value of $\epsilon/15$ is, but for ethylene carbonate, due to its high dielectric constant, $a_w = 3$ is preferred. Formamide, which has the highest ϵ value, has $a_w = 7$, and the contribution due to the hydrogen bond is neglected, while for morpholine, with its very low ϵ value, $a_w = 1$ has been preferred. Values of $a_w \neq 1$ are collected in Table 4 first column (in parentheses). This rather rough "ad hoc noncovalent" parameter shows the following descriptive power for the dielectric constant: $Q = 0.105$, $F = 250.3$, $r = 0.895$, $s = 8.49$, and $n = 64$. This nonperfect description of the dielectric constant renders the a_w index different from the dielectric constant.

Let us start with the most studied modeling, the modeling of the boiling points, T_b (or BP) of $n = 63$ solvents. The modelings due to the molar M masses, the dielectric ϵ constant, and to a_w are very bad with

$$M\colon\ Q = 0.006,\ F = 7.9 \qquad \epsilon\colon\ Q = 0.09,\ F = 16.4 \qquad a_w\colon\ Q = 0.08,\ F = 14.2$$

If these 'ad hoc' descriptors have to play a role in the modeling of this property, then from these ratings they should be ϵ and/or a_w, mainly. The best molecular connectivity model is achieved by a δ^v-based valence molecular connectivity index: $\{^0\chi^v\}$ with $Q = 0.013$, $F = 33.9$, $r = 0.598$, and $s = 36.4$. Neither the use of linear combination of molecular connectivity indices (LCCI) nor the use of molecular connectivity indices based on the δ^z cardinal number, and their linear combinations, improve the modeling. Introduction, instead, of the following set of semiempirical supramolecular connectivity indices improves the description quite a lot:

$$\{a_w(\chi,\chi^v)\} = \{a_wD,\ a_wD^v,\ a_w{}^0\chi,\ a_w{}^0\chi^v,\ a_w{}^1\chi,\ a_w{}^1\chi^v,\ \chi_t/a_w,\ \chi_t{}^v/a_w\} \quad (9)$$

and the corresponding supraindex set determined by indices of

TABLE 3: Molecular Connectivity Indices, with Valence $^i\chi^v$ Indices Based on Cardinal δ^v Number, for Compounds Given in Table 1

solvent	D	D^v	$^0\chi$	$^0\chi^v$	$^1\chi$	$^1\chi^v$	χ_t	χ_t^v
acetone	6	12.0	3.577 35	2.908 25	1.732 05	1.204 12	0.577 35	0.204 12
acetonitrile	4	10.0	2.707 11	1.947 21	1.414 21	0.723 61	0.707 11	0.223 61
benzene	12	18	4.242 64	3.464 10	3	2	0.125 00	0.037 04
benzonitrile	16	28	5.819 99	4.333 97	3.931 85	2.384 29	0.072 17	0.007 17
1-butanol	8	12	4.121 32	3.568 53	2.414 21	2.023 33	0.353 55	0.158 11
2-butanone	8	14	4.284 46	3.615 36	2.270 06	1.764 78	0.408 25	0.144 34
butyl acetate	14	24	6.405 78	5.437 82	3.770 06	2.904 03	0.144 34	0.029 46
CS_2	4	5.33333	2.707 11	2.949 49	1.414 21	1.224 74	0.707 11	0.750 00
CCl_4	8	7.11112	4.500 00	5.035 57	2.000 00	2.267 78	0.500 00	0.826 53
Cl−benzene	14	19.7778	5.112 88	4.520 64	3.393 85	2.477 63	0.102 06	0.036 37
1Cl−butane	8	7.777 78	4.121 32	4.255 21	2.414 21	2.508 89	0.353 55	0.400 89
$CHCl_3$	6	5.333 34	3.577 35	3.979 03	1.732 05	1.963 96	0.577 35	0.841 69
cyclohexane	12	12	4.242 64	4.242 64	3	3	0.125 00	0.125 00
cyclopentane	10	10	3.535 53	3.535 53	2	2	0.176 78	0.176 78
1,2-dCl−benzene	16	21.5556	5.983 13	5.577 18	3.804 53	2.961 24	0.083 33	0.035 71
1,2-dCl-ethane	6	5.555 56	3.414 21	3.682 00	1.914 21	2.103 57	0.500 00	0.642 86
dCl-methane	4	3.555 56	2.707 11	2.974 89	1.414 21	1.603 57	0.707 11	0.909 13
N,*N*-dM-acetamd	10	18	5.154 70	4.355 46	2.642 73	1.822 16	0.333 33	0.091 29
N,*N*-dM-formamd	8	16	4.284 46	3.432 81	2.270 06	1.388 33	0.408 25	0.105 41
1,4-dioxane	12	20	4.242 64	3.644 92	3	2.154 70	0.125 00	0.041 67
ether	8	12	4.121 32	3.822 46	2.414 21	1.991 56	0.353 55	0.204 12
ethyl acetate	10	20	4.991 56	4.023 60	2.770 06	1.904 03	0.288 68	0.058 93
ethyl alcohol	4	8	2.707 11	2.154 32	1.414 21	1.023 33	0.707 11	0.316 23
heptane	12	12	5.535 53	5.535 53	3.414 21	3.414 21	0.176 78	0.176 78
hexane	10	10	4.828 43	4.828 43	2.914 21	2.914 21	0.250 00	0.250 00
2-methoxyethanol	8	16	4.121 32	3.269 68	2.414 21	1.513 15	0.353 55	0.091 29
methyl alcohol	2	6	2	1.447 21	1	0.447 21	1	0.447 21
2-methylbutane	8	8	4.284 46	4.284 46	2.270 06	2.270 06	0.408 25	0.408 25
4-M-2-pentanone	12	18	5.861 81	5.192 71	3.125 90	2.620 63	0.235 70	0.083 33
2-M-1-propanol	8	12	4.284 46	3.731 67	2.270 06	1.879 18	0.408 25	0.182 57
2-M-2-propanol	8	12	4.5	3.947 21	2	1.723 61	0.5	0.223 61
DMSO	6	8.666 67	3.577 35	3.632 99	1.732 05	2.949 48	0.577 35	0.5
nitromethane	6	18	3.577 35	2.263 71	1.732 05	0.812 36	0.577 35	0.074 54
1-octanol	16	20	6.949 75	6.396 96	4.414 21	4.023 33	0.088 39	0.039 53
pentane	8	8	4.121 32	4.121 32	2.414 21	2.414 21	0.353 55	0.353 55
3-pentanone	10	16	4.991 56	4.322 46	2.808 06	2.325 44	0.288 68	0.102 06
1-propanol	6	10	3.414 21	2.861 43	1.914 21	1.523 33	0.5	0.223 61
2-propanol	6	10	3.577 35	3.024 56	1.732 05	1.412 90	0.577 35	0.258 20
pyridine	12	20	4.242 64	3.333 97	3	1.849 73	0.125 00	0.028 69
ttCl-ethylene	10	11.1111	5.154 70	5.535 57	2.642 73	2.517 78	0.333 33	0.413 26
tt-hydrofuran	10	14	3.535 53	3.236 68	2.5	2.077 35	0.176 78	0.102 06
toluene	14	20	5.112 88	4.386 75	3.393 85	2.410 68	0.102 06	0.032 08
1,1,2tCltFEthane	14	31.3333	7	5.535 57	3.25	2.517 78	0.25	0.019 68
2,2,4-tM-pentane	14	14	6.784 46	6.784 46	3.416 50	3.416 50	0.204 12	0.204 12
o-xylene	16	22	5.983 13	5.309 40	3.804 53	2.827 35	0.083 33	0.027 78
p-xylene	16	22	5.983 13	5.309 40	3.787 69	2.821 37	0.083 33	0.027 78
acetic acid	6	16	3.577 35	2.355 46	1.732 05	0.927 73	0.577 35	0.091 29
decaline	22	22	6.811 55	6.811 55	4.966 33	4.966 33	0.020 83	0.020 83
dBr-methane	4	2.518 52	2.707 11	4.635 02	1.414 21	2.777 46	0.707 11	2.727 40
1,2-dCl-E-enZ	6	7.555 56	3.414 21	3.422 48	1.914 21	1.642 64	0.5	0.428 57
1,2-dCl-E-enE	6	7.555 56	3.414 21	3.422 48	1.914 21	1.642 64	0.5	0.428 57
1,1-dCl-E-en	6	7.555 56	3.577 35	3.474 89	1.732 05	1.487 45	0.577 35	0.454 57
dMethoxymethane	8	16	4.121 32	3.523 60	2.414 21	1.393 85	0.353 55	0.117 85
dMethyl ether	4	8	2.707 11	2.408 25	1.414 21	0.816 50	0.707 11	0.408 25
E-enCarbonate	10	26	4.991 56	3.048 17	2.770 06	1.275 81	0.288 68	0.015 21
formamide	4	12	2.707 11	1.562 95	1.414 21	0.569 04	0.707 11	0.136 08
methyl chloride	2	1.777 78	2	2.133 89	1	1.133 89	1	1.3389
morpholine	12	18	4.242 64	3.736 67	3	2.284 46	0.125	0.051 03
quinoline	22	34	6.811 55	5.488 67	4.966 33	3.264 50	0.020 83	0.002 39
SO_2	4	13.6667	2.707 11	1.591 09	1.414 21	0.632 45	0.707 11	0.129 10
2,2-ttCl-ethane	10	9.111 12	5.154 70	5.690 27	2.642 73	2.951 94	0.333 33	0.551 02
ttM-urea	14	24	6.732 05	5.802 68	3.553 42	2.440 19	0.192 45	0.040 82
tCl-E-en	8	9.333 34	4.284 46	4.479 03	2.270 06	2.077 22	0.408 25	0.420 85

eq 2. Why total χ_t and χ_t^v have been divided by a_w becomes clear considering the definitions of these indices (see eq 6); the bigger the molecule the smaller the total χ indices, the contrary being true for the other indices of the set. The improvement in the modeling of T_b is noticeable either at the level of the single supraindex as well as at the level of a multi-supraindex description,

$$\{a_wD\}: Q = 0.030, F = 182.9, r = 0.866, s = 29.0$$

$$\{a_wD, \chi_t^v/a_w\}: Q = 0.031, F = 98.0, r = 0.875, s = 28.3$$

$$\{a_w{}^1\chi, \chi_t/a_w, \chi_t^v/a_w\}: Q = 0.033, F = 73.7, r = 0.888, s = 27.4$$

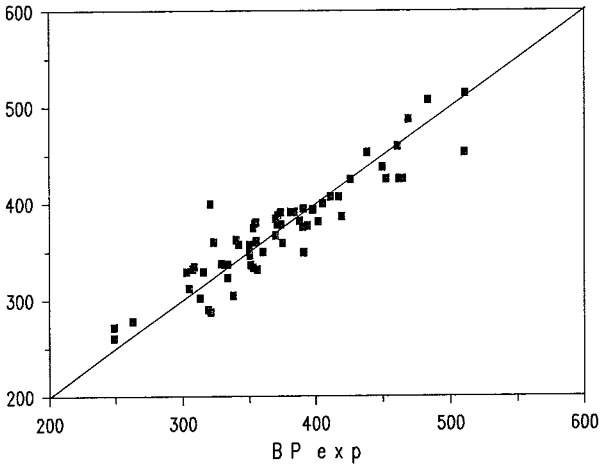

Figure 1. Plot of the calculated versus the experimental boiling points, BP, of 63 organic solvents.

This description can further be improved with the introduction of the following semiempirical molecular connectivity term found with a trial-and-error procedure,

$$X_{BP} = (a_w D)^{0.7}/(\chi_t/a_w + 1.5) \qquad (10)$$

which shows a $Q = 0.031$, $F = 198$, $r = 0.87$, $s = 28.1$, and $\langle u \rangle = 22$. The improvement relative to the single supraindex is not impressive, but now, the following enhanced description with linear combinations of X_{BP} and $\{a_w\chi\}$ supraindices, where X_{BP} is a dominant descriptor (in this case the total combinatorial search goes over to a forward selection search[2,3]), can be obtained

$$\{X_{BP}, \chi_t^v/a_w\}: Q = 0.034, F = 119.6, r = 0.894, s = 26.2, \langle u \rangle = 13.0$$

$$\{X_{BP}, \chi_t^v/a_w, \chi_t/a_w\}: Q = 0.035, F = 85.5, r = 0.902, s = 25.5, \langle u \rangle = 6.60$$

$$\{X_{BP}, \chi_t^v/a_w, \chi_t/a_w, \epsilon, M\}: Q = 0.039, F = 63.1, r = 0.920, s = 23.4, \langle u \rangle = 4.63$$

The last description, even if its F and $\langle u \rangle$ have worsened along the modeling, can be used to model the boiling points of the given solvents. The vectors used to model the boiling points of Figure 1 are (where $X^0 \equiv 1$):

$$X = (X_{BP}, \chi_t^v/a_w, \chi_t/a_w, \epsilon, M, X^0)$$

$$\mathbf{C} = (21.6930, 30.2553, -96.2886, 0.72952, 0.31665, 283.385)$$

$$\mathbf{u} = (4.9, 1.8, 2.7, 3.3, 2.8, 12.3)$$

TABLE 4: Molecular Connectivity Valence Indices $^i\chi^z$ Based on δ^z Cardinal Number (Normal $^i\chi$ Indices Equal to $^i\chi$ Indices of Table 2), for Compounds Given in Table 1, with Heteroatoms Other Than Carbons (In parentheses, near the solvent name, values for $a_w \neq 1$ (see text))

solvent (a_w)	D^z	$^0\chi^z$	$^1\chi^z$	χ_t^z
acetone (2)	9	3.077 35	1.288 68	0.288 68
acetonitrile (2.5)	7.5	2.132 46	0.816 23	0.316 23
benzonitrile (1.5)	25.5	4.519 21	2.476 91	0.010 14
1-butanol (2)	9.5	3.753 78	2.154 32	0.223 61
2-butanone	11	3.784 46	1.849 34	0.204 12
butyl acetate	18	5.776 02	3.192 71	0.058 93
CS_2	8	1.914 21	0.707 11	0.25
CCl_4	13.3333	3.118 62	1.309 31	0.091 84
Cl-benzene	21.3333	4.041 41	2.238 01	0.021 00
1Cl-butane	9.33333	3.775 97	2.170 02	0.231 46
$CHCl_3$	3.99999	2.541 31	1.133 89	0.161 99
1,2-dCl-benzene	24.6666	4.618 71	2.482 00	0.011 90
1,2-dCl-ethane	8.66666	2.723 52	1.425 82	0.214 29
DCl-methane	6.66666	2.016 42	0.925 82	0.303 05
N,*N*-dM-acetamd (2.5)	12.5	4.709 81	2.369 81	0.182 57
N,*N*-dM-formamd (2.5)	10.5	3.787 16	1.963 39	0.210 82
1,4-dioxane	14	3.983 13	2.632 99	0.083 33
ether	9	3.991 56	2.230 71	0.288 68
ethyl acetate	14	4.361 81	2.192 71	0.117 85
ethyl alcohol (2)	5.5	2.339 56	1.154 32	0.447 21
2-methoxyethanol (2)	10.5	3.624 02	1.932 81	0.082 57
methyl alcohol (2)	19	5.809 40	2.898 72	0.055 56
2-methylbutane	3.5	1.632 46	0.632 46	0.632 46
4-M-2-pentanone	15	5.361 81	2.705 18	0.117 85
2-M-1-propanol (2)	9.5	3.916 91	2.010 16	0.258 20
2-M-2-propanol (2)	9.5	4.132 46	1.816 23	0.316 23
DMSO (3)	10	3.115 36	1.702 89	0.288 68
nitromethane (2.5)	9.5	2.787 16	1.362 75	0.210 82
1-octanol (2)	17.5	6.582 20	4.154 32	0.055 90
3-pentanone	13	4.491 56	2.410 00	0.144 34
1-propanol (2)	7.5	3.046 67	1.654 32	0.316 23
2-propanol (2)	7.5	3.209 81	1.519 85	0.365 15
pyridine	17.5	3.519 21	2.063 63	0.040 57
TtCl-ethylene	17.3333	3.618 62	1.559 31	0.045 92
Tt-hydrofuran	11	3.405 78	2.316 50	0.144 34
1,1,2tCltFEthane	25.4999	4.567 53	2.033 76	0.010 71
acetic acid (2)	10.5	2.709 81	1.104 90	0.182 57
DBr-methane	5.5	2.218 96	1.069 04	0.404 06
1,2-dCl-E-enZ	10.6666	2.464 01	1.089 26	0.142 86
1,2-dCl-E-enE	10.6666	2.464 01	1.089 26	0.142 86
1,1-dCl-E-en	10.6666	2.516 42	1.008 21	0.151 52
DMethoxymethane	12	3.707 11	1.707 11	0.176 78
DMethyl ether	6	2.5	1	0.5
E-enCarbonate (3)	18.5	3.494 26	1.551 83	0.037 27
formamide (7)	7.5	1.971 20	0.804 74	0.272 17
methyl chloride	3.33333	1.654 65	0.654 65	0.654 65
morpholine	13	4.112 88	2.816 50	0.102 06
quinoline	31.5	5.673 91	3.464 07	0.003 38
SO_2	8	1.861 81	0.816 50	0.235 70
2,2-ttCl-ethane	15.3333	3.773 32	1.845 19	0.061 22
TtM-urea (1.5)	16	6.342 26	3.450 95	0.115 47
TCl-E-en	13.9999	3.041 31	1.321 29	0.080 99

Utilities of the single indices, especially of the first (up to 17, and $\langle u \rangle$ =11), second and last one, can further be improved with the introduction of orthogonal indices, as has been shown elsewhere.[3,5−7,35] The kind of descriptors involved in the modeling allow us to understand the structural basis of this property. While X_{BP} is made up of nonvalence molecular connectivity indices, which are shape-dependent, the improvement caused with the introduction of the total valence supraindex χ_t^v/a_w underlines the importance of a pseudograph representation for these molecules. At the same time, the indirect influence of a_w and further of ϵ, shows that also subtle electrostatic intermolecular interactions are important for T_b (van der Waals and hydrogen bond type of interactions). The additional improvement caused by the molar mass M (without M we have $Q = 0.037$, $F = 69$, and $r = 0.909$) seems to tell us that bulk factors contribute in some minor way to T_b. Recent and less recent studies on boiling points[46,47] of some nonpolar or slightly polar compounds have underlined the importance of polarizability and further of shape and size (molar volume) on T_b. Molecular weight here can indirectly help, through density, to model the size factor, which determines the boiling points. Clearly, a better accuracy can be achieved and has been achieved with more homogeneous classes of compounds and with different molecular structure indices, like the recent modeling of C_2−C_{10} alkenes and cycloalkenes,[48] with C_1−C_2 and C_1−C_4 chlorofluoroalkenes[49] and C_2 chlorofluoroalkenes,[50] while a modeling based on a linear combination of more sophisticated nontopological indices achieved recently a very interesting modeling of a wide set of boiling points and of critical transition temperatures.[51-53] Our aim, nevertheless, is to achieve a modeling in a rather *pedestrian* way, i.e., to maintain the calculations at an elementary level both formally and practically.

Normally, the modeling of melting points, T_m (or MP), has always been an awkward task,[3,5] as they seem to have an "all-or-none" character at the level of each compound. The modeling can, in fact, be achieved only by the aid of a kind of "*double sieve*" mechanism, where, after an approximate modeling with X terms has been achieved, a second more individual modeling is to be searched 'a posteriori' at the level of each small subset of compounds, in a kind of group equivalent scheme, where each subset seems to be influenced by its own steric and noncovalent interactions. This means that a satisfactory structure-related descriptor for melting points still hides somewhere. But let us anyway try to derive by the aid of this 'double sieve' mechanism a satisfactory T_m modeling for $n = 62$ compounds. Modeling by M, ϵ, and a_w is extremely awkward with $Q(M) = 0.007$, $F(M) = 5.7$; $Q(\epsilon) = 0.007$, $F(\epsilon) = 5.8$; and $Q(a_w) = 0.006$, $F(a_w) = 5.1$, and it is very difficult at this level to understand which of these three descriptors will play some role in the modeling of T_m. With linear combinations of molecular connectivity indices, together with the dielectric constant, things improve but in a still unsatisfactory way,

$$\{D^v\}: Q = 0.010, F = 14.0, r = 0.435, s = 42.6$$

$$\{D, \epsilon\}: Q = 0.014, F = 13.8, r = 0.564, s = 39.4$$

$$\{D, \epsilon, {}^1\chi\}: Q = 0.017, F = 12.2, r = 0.622, s = 37.7$$

and here δ^z-based χ indices, together with the intermolecular a_w parameter, fare better

$$\{\chi_t^z\}: Q = 0.011, F = 17.7, r = 0.477, s = 41.6$$

$$\{\chi_t^z, a_w\}: Q = 0.016, F = 17.6, r = 0.611, s = 37.7$$

$$\{\chi_t^z, a_w, D^z\}: Q = 0.017, F = 13.4, r = 0.640, s = 36.9$$

In both cases introduction of M improves the modeling a little bit ($Q = 0.018$, $F = 10.5$, $r = 0.651$, for the δ^z-based χ indices), while linear combinations with more than these four descriptors fare worse. Use of supramolecular connectivity indices $\{a_w(\chi,\chi^z)\}$ of eq 9 where χ^z-type indices are used, improved somehow the modeling, especially at the level of the single descriptor

$$\{\chi_t^z/a_w\}: Q = 0.015, F = 29.5, r = 0.574, s = 38.7$$

A further small improvement can be reached with the following semiempirical molecular connectivity term based on a δ^z-type of χ index, while its linear combinations with other $\{\chi, \chi^z\}$ or $\{a_w(\chi,\chi^z)\}$ indices do not ameliorate the modeling in a consistent way

$$X_{MP} = a_w D/(a_w{}^1\chi + 0.8{}^1\chi^z) \qquad (11)$$

$$\{X_{MP}\}: Q = 0.018, F = 41.7, r = 0.640, s = 36.3$$

To really improve the modeling in a consistent way, use has to be made of the second step of a 'double sieve' mechanism (as already done for amino acids[5]) to further differentiate the set of compounds while considering what has already been done with X_{MP} as the first step of the 'double sieve'. Index a_w will be rewritten for subsets of these compounds. It will be assumed that for C_6 rings (benzene, cyclohexane, morpholine, and dioxane) $a_w = 2$, for C_5 rings (cyclopentane and THF) $a_w = 0.5$, for the couple *o*- and *p*-xylene $a_w(o) = 1$ and $a_w(p) = 2$, for alcohols $a_w = 1$, but for the long chain of *n*-octanol $a_w = 3$, while for CH_4-like derivatives we assume the following: methanol and MeCl, $a_w = 2$, 2-M-2-propanol and dBr-Me, $a_w = 4$. For urea, which is rather similar to *N*,*N*-DMFA and *N*,*N*-DMAM, we assume $a_w = 2.5$, and for acetic acid, $a_w = 4$, which nearly doubles the value of acetonitrile. For the series ether, dMEther and acetone it will be assumed that $a_w = 0.7$, while for 3-pentanone, $a_w = 1.4$. For the series DMSO, SO_2, and CS_2, where the ratios of atoms other than S over S are 9, 2, and 0.5, we can define $a_w = 9, 2, 0.5$, respectively, and for E-enCarbonate we simply adopt the $\epsilon/15$ value of 6. Finally, for the series 1,2-dCl-E-enZ, 1,2-dCl-E-enE, and 1,1-dCl-E-en we adopt, $a_w = 2$, 1, and 0.5, respectively. With this new 'ad hoc' parametrization the modeling of T_m improves to

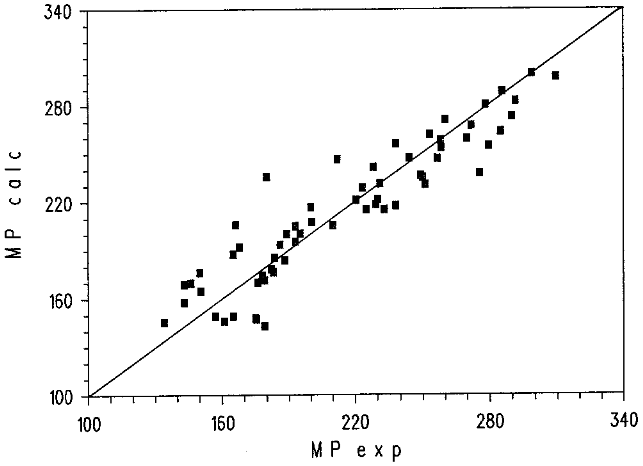

Figure 2. Plot of the calculated vs experimental melting points, MP, of 62 organic solvents.

$$\{X_{MP}\}: Q = 0.051, F = 353, r = 0.925, s = 18.0, \quad \langle u \rangle = 10.0$$

Calculated T_m values (in K) of Figure 2 can be obtained with the following vectors

$$\mathbf{X} = (X_{MP}, X^0), \qquad \mathbf{C} = (84.9093, 12.6086), \quad \mathbf{u} = (18.8, 1.2)$$

where the constant term has a rather low utility, which nevertheless improves consistently relative to the former modeling, which had $\langle u \rangle = 3.6$ and $\mathbf{u} = (6.5, 0.8)$. Such a 'double sieve' mechanism can be conducted in many different ways, and on the basis of different assumptions, our aim here was primarily to show how a difficult modeling can be improved with further assumptions that can be deduced from an initial unsatisfactory modeling. At the level of a 'double sieve' approximation, a clear distinction between the ratings of δ^v- or δ^z-based χ indices is rather Byzantine. Now, to further check the 'double sieve' modeling mechanism, let us test which of the two terms, X_{BP} or X_{MP}, is able to model also the other temperature, expanding thus the modeling capability of one of the found terms. While X_{MP} is a bad descriptor of T_b, X_{BP} is not an insignificant descriptor of T_m, with $Q = 0.031$ and $F = 130.2$, and it can even be improved, optimizing its constant parameters: in fact, the following term

$$X_{\mathrm{MP,BP}} = (a_\mathrm{w}D)^{0.2}/[(\chi_\mathrm{t}/a_\mathrm{w})^{0.2} + 15] \tag{12}$$

for the melting points T_m rates $Q = 0.035$, $F = 160.4$, $r = 0.85$, $s = 24.7$, $\langle u\rangle = 7.3$, and for the boiling points T_b, $Q = 0.029$, $F = 175.7$, $r = 0.86$, $s = 29.4$, $\langle u\rangle = 7.5$. Thus a single semiempirical term could be used to approximately model both temperatures, for a total of 125 values. The major drawback of this descriptor is that it is a dead-end descriptor; i.e., its descriptive power cannot be improved with the help of any linear combination with common χ-type indices.

Even the refractive index, RI or n_r, which should not be mistaken with molar refractivity, R_m, which is a factor in determining the boiling points of molecular compounds[46,47] and which is determined by n_r through the equation $R_\mathrm{m} = M(n_\mathrm{r}^2 - 1)/d(n_\mathrm{r}^2 + 1)$ (M, molar mass, d = density), is better modeled by a δ^v-based set of molecular connectivity indices given by eq 1. The best descriptor for this property for $n = 61$ compounds is $\{\chi_\mathrm{t}^\mathrm{v}\}$ with $Q = 6.00$, $F = 49.6$, $r = 0.676$, and $s = 0.11$, while M, ϵ, and a_w show the following descriptive power, respectively: $Q(M) = 3.52$, $F(M) = 17.0$; $Q(\epsilon) = 0.47$, $F(\epsilon) = 0.31$; $Q(a_\mathrm{w}) = 0.68$, $F(a_\mathrm{w}) = 0.63$. From these values it should be expected that M should surely play some role in the description of RI. The following LCCI, where $\chi_\mathrm{t}^\mathrm{v}$ is a dominant index, offers a better description for RI,

$$\{\chi_\mathrm{t}^\mathrm{v}, D^\mathrm{v}\}: Q = 10.2, F = 71.7, r = 0.844, s = 0.08$$

$$\{\chi_\mathrm{t}^\mathrm{v}, \chi_\mathrm{t}, D^\mathrm{v}, {}^0\chi, a_\mathrm{w}\}: Q = 14.8, F = 60.1, r = 0.919, s = 0.06$$

But the following dominant X_RI molecular connectivity term alone can explain most of the modeling

$$\{X_\mathrm{RI} = (\chi_\mathrm{t}^\mathrm{v})^3/(\chi_\mathrm{t})^{2.5}\}: Q = 13.6, F = 256, r = 0.902, s = 0.07, \langle u\rangle = 90.4$$

a modeling that can be improved with the following linear combinations, where the second five-index combination includes the empirical parameters M and a_w

$$\{X_\mathrm{RI}, D\}: Q = 16.7, F = 198, r = 0.932, s = 0.06, \langle u\rangle = 33.3$$

$$\{X_\mathrm{RI}, D, {}^0\chi, M, a_\mathrm{w}\}: Q = 19.9, F = 109, r = 0.953, s = 0.05, \langle u\rangle = 11.1$$

Eliminating from the modeling two strong outliers (CS_2 and decalin), i.e., with $n = 59$, the description improves to

$$\{X_\mathrm{RI}\}: Q = 14.7, F = 290, r = 0.914, s = 0.06, \langle u\rangle = 94.3$$

$$\{X_\mathrm{RI}, D\}: Q = 20.8, F = 289, r = 0.955, s = 0.05, \langle u\rangle = 38.2$$

$$\{X_\mathrm{RI}, D, {}^0\chi, a_\mathrm{w}, M\}: Q = 30.0, F = 242, r = 0.979, s = 0.03, \langle u\rangle = 16.3$$

Relative to the significant improvement in Q the small decrease in F can be neglected, and the overall utility even if decreasing continues to be meaningful. A total of 61 calculated RI values can be obtained with the following vectors (notice the good utility values of the parameters)

$$\mathbf{X} = (X_\mathrm{RI}, D, {}^0\chi, a_\mathrm{w}, M, X^0)$$

$$\mathbf{C} = (0.02135, 0.02478, -0.06271, 0.02591, 0.00103, 1.32902)$$

$$\mathbf{u} = (24.1, 10.0, 6.7, 5.0, 4.8, 60.4)$$

In this description, the a_w descriptor appears directly in the description, and two molecular connectivity total indices build up the connectivity term; further, two nonvalence D and ${}^0\chi$ indices help to improve the description. Even here (see T_b) M contributes a little bit to the description; in fact, without M the description is $Q = 25.1$, $F = 212$, $r = 0.970$. Intermolecular interactions seem to play here a lesser role than in boiling point modeling, actually up to the rather good description with three indices, $\{X_\mathrm{RI}, D, {}^0\chi\}$ with $Q = 23.3$, $F = 242$, $r = 0.964$ for $n = 59$ and $Q = 17.9$, $F = 147$, $r = 0.941$, $s = 0.05$ for $n = 61$, no intermolecular a_w or bulk M descriptors are required.

In modeling the densities, d, of $n = 62$ solvents δ^v- and δ^z-based χ indices show rather similar ratings,

$$\{\chi_\mathrm{t}^\mathrm{v}\}: Q = 1.64, F = 12.2, r = 0.41, s = 0.25$$

and

$$\{{}^1\chi^\mathrm{z}\}: Q = 1.70, F = 13.1, r = 0.42, s = 0.25$$

Using linear combinations of these indices, a hardly noticeable improvement in Q can be detected, while descriptors M, ϵ, and a_w rate as follows: $Q(M) = 2.75$, $F(M) = 34.1$; $Q(\epsilon) = 0.21$, $F(\epsilon) = 0.20$; $Q(a_\mathrm{w}) = 0.16$, $F(a_\mathrm{w}) = 0.11$: I.e., descriptor M rates better than $\chi_\mathrm{t}^\mathrm{v}$ and ${}^1\chi^\mathrm{z}$, and it should be expected that it will play some interesting role in the description of this property. The importance of M is not unexpected, as densities are strictly related to molar masses. In fact, the following linear combination of a molecular connectivity index and M show a remarkably improved modeling

$$\{{}^0\chi^\mathrm{v}, M\}: Q = 9.71, F = 213.0, r = 0.937, s = 0.10$$

which can be further improved by introducing the following set of molar-mass-based χ indices (where total indices are multiplied, as total indices describe an inverted domain relative to the other indices),

$$\{M^{-1}(\chi,\chi^\mathrm{v})\} = \{D/M, D^\mathrm{v}/M, {}^0\chi/M, {}^0\chi^\mathrm{v}/M, {}^1\chi/M, {}^1\chi^\mathrm{v}/M, \chi_\mathrm{t}M, \chi_\mathrm{t}^\mathrm{v}M\} \tag{13}$$

These indices offer a significant single-index and multi-index modeling, the single index being a leading descriptor that transforms the complete combinatorial search into a much easier forward selection search

$$\{{}^0\chi^\mathrm{v}/M\}: Q = 8.88, F = 356.8, r = 0.925, s = 0.10$$

$$\{{}^0\chi^\mathrm{v}/M, {}^1\chi/M\}: Q = 10.9, F = 269.6, r = 0.949, s = 0.09$$

$$\{{}^0\chi^\mathrm{v}/\mathrm{M}, {}^1\chi/M, D/M\}: Q = 15.9, F = 379.8, r = 0.975, s = 0.06$$

With four descriptors the improvement stagnates and then starts to decrease; δ^z-based molar-mass χ indices offer only a good single-index description

$$\{{}^0\chi^\mathrm{z}/M\}: Q = 9.11, F = 375.2, r = 0.929, s = 0.10$$

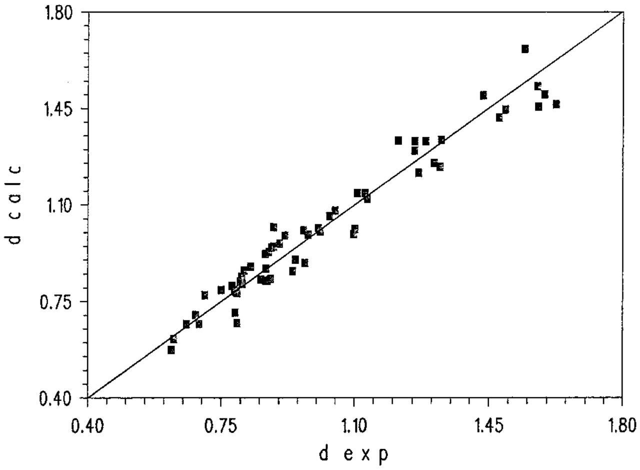

Figure 3. Plot of the calculated vs experimental density, *d*, values of 62 organic solvents.

The best overall description is achieved with the following semiempirical molecular connectivity term, derived by a trial-and-error procedure centered around ${}^0\chi^v$ and M. This term is a dead-end descriptor, as no further improvement can be found with a linear combination of this and other indices of sets of eqs 1 or 13

$$X_d = {}^0\chi^v({}^1\chi + \chi_t)^{0.4}/M^{1.4} \qquad (14)$$

$$\{X_d\}: Q = 15.6, F = 1094.7, r = 0.974, s = 0.06, \langle u\rangle = 47.4$$

A total of 62 calculated values of Figure 3 have been obtained with the following vectors of the single X_d term, where $X^0 \equiv 1$

$$\mathbf{X} = (X_d, X^0), \qquad \mathbf{C} = (-96.5416, 2.11059), \qquad \mathbf{u} = (33.1, 61.7)$$

The negative value of the slope for X_d underlines the fact the density and X_d are inversely related; that is, the numerator of X_d encodes the volume attribute of the property. The term given by eq 14 is a special case of a more convoluted semiempirical term given by eq 15

$$X_d(0.12) = [{}^0\chi^v({}^1\chi + \chi_t)^{0.4}]/[\mathrm{M}^{1.4} + (0.12\epsilon)^{1.2}] \qquad (15)$$

which rates even better than term 14, with $Q = 16.1$, $F = 1167.8$, and $r = 0.975$. But the small loss in quality due to term 14 is more than compensated by its higher simplicity. Before closing this modeling let us notice that around ${}^0\chi^v$ is also centered a term that perfectly models the side-chain volume of amino acids.[3]

The modeling of flashpoints, FP, of $n = 41$ solvents with M, ϵ, and a_w, whose rating are $Q(M) = 0.01$, $F(M) = 4.9$; $Q(\epsilon) = 0.02$, $F(\epsilon) = 17.8$; and $Q(a_w) = 0.02$, $F(a_w) = 21.6$, reveals that two of them, at least, should play some role all along the FP modeling. While at the level of sets of eq 1 and of eq 2, the set of eq 1 fares better with $\{D^v\}$, $Q = 0.015$, $F = 11.2$ and $r = 0.47$, while $\{D^z\}$ fares $Q = 0.012$, $F = 7.0$, $r = 0.39$. As linear combinations of χ indices of set (1) do not improve the modeling, we resort to indices of eq 9, i.e., $\{a_w(\chi,\chi^v)\}$ and the corresponding $\{a_w(\chi,\chi^z)\}$. Here, supradescriptors $\{a_w(\chi,\chi^z)\}$ fare better, even at the level of a multi-supraindex description

$$\{a_wD^z\}: Q = 0.049, F = 117.6, r = 0.867, s = 17.6$$

$$\{a_wD^z, \epsilon, M\}: Q = 0.069, F = 79.9, r = 0.928, s = 13.5$$

$$\{a_wD^z, \epsilon, M, a_w\}: Q = 0.073, F = 64.2, r = 0.937, s = 12.9$$

In fact, the best $\{a_w(\chi,\chi^v)\}$ supraindex description is $\{a_wD^v, a_w{}^1\chi^v, M\}$, which fares $Q = 0.047$, $F = 35.8$, $r = 0.862$. The last $\{a_wD^z, \epsilon, M, a_w\}$ description has not only a rather low F value but the overall utility as well as the single utilities is rather unsatisfactory: $\langle u\rangle = 4.2$, u = (2.9, 3.3, 4.7, 2.1, 7.8). Clearly, we could resort to the corresponding supraorthogonal indices, which normally have better utilities, and sometimes even improved single-descriptor power, but let us see if a semiempirical term derived from the four descriptors $\{a_wD^z, \epsilon, M, a_w\}$ fares better. In fact, the following dead-end term can be used to model FP

$$X_{FP} = (a_wD^z)^{0.2}(\epsilon + 0.4M)^{0.4} \qquad (16)$$

This term has $Q = 0.073$, $F = 258$, $r = 0.932$, $s = 12.8$, $\langle u\rangle = 15.4$, where also the single utilities are very interesting, as can be seen from vectors used to obtain the 41 calculated values (as some FP values are negative, eq 7 has to be used without operational bars)

$$\mathbf{X} = (X_{FP}, X^0), \qquad \mathbf{C} = (22.8502, -171.989), \qquad \mathbf{u} = (16.1, 14.8)$$

After the nice modeling of density and flash points, the modeling of viscosity, η (or VIS), of $n = 38$ solvents was expected to be rather straightforward. Instead, this modeling happened to be rather difficult. Things start to be disappointing with M, ϵ, and a_w, with $Q(M) = 0.10$, $F(M) = 0.14$; $Q(\epsilon) = 0.54$, $F(\epsilon) = 4.5$, $Q(a_w) = 1.1$, $F(a_w) = 19.6$. Molecular connectivity δ^v-based indices (set of eq 1) are better descriptors than the corresponding δ^z-based ones, but the description is, nevertheless, very poor; for example, the best single index, $\{{}^0\chi^v\}$, has $Q = 0.17$, $F = 0.43$, $r = 0.11$. The modeling shows a consistent improvement with the introduction of supramolecular connectivity indices of set 9, in fact the best single index is $\{a_w{}^0\chi^v\}$: $Q = 1.11$, $F = 19$, $r = 0.59$, indeed, a result similar to the score of a_w alone. Now, redefining a_w for alcohols in the following way, $a_w(\mathrm{R{-}OH}) = 2 + \epsilon/15 = a_{OH}$ (truncated at the second figure), we can achieve a further improvement with the following single leading, and two descriptor, terms

$$\{a_{OH}{}^0\chi^v\}: Q = 2.37, F = 43.2, r = 0.845, s = 0.36$$

$$\{a_{OH}{}^0\chi^v, a_{OH}{}^1\chi^v\}: Q = 2.46, F = 31.1, r = 0.856, s = 0.35$$

But after these two descriptions supraindices are unable to improve the modeling. Meanwhile, let us notice that $Q(a_{OH}) = 1.13$ and $F(a_{OH}) = 19.6$; that is, description by a_{OH} alone cannot be compared to the two previous descriptions. A trial-and-error search for a good term ends up with the following crucial semiempirical X_{VIS} term, which rates $Q = 3.21$, $F = 158.8$, $r = 0.903$, $s = 0.28$, $\langle u\rangle = 6.3$

$$X_{VIS} = (a_{OH}MD)^{1.7}/(1.6\,{}^0\chi + 3)^{4.3} \qquad (17)$$

This term, which is independent of δ^v- or δ^z-based χ indices, depending only on the connectivity structure of a chemical graph, is influenced not only by a_{OH} but also by M, a not unexpected result for viscosity. To model this property, the following combination has been chosen (combinations with more indices are poorer descriptors)

$$\{X_{\mathrm{VIS}}, a_{\mathrm{OH}}{}^{1}\chi^{\mathrm{v}}\}: Q = 3.30, F = 83.9, r = 0.910, s = 0.28, \langle u \rangle = 2.7$$

The average utility of this combination is rather low, due to the low utility of the second and third constant component of the correlation vectors used to model the η values,

$$\mathbf{X} = (X_{\mathrm{VIS}}, a_{\mathrm{OH}}{}^{1}\chi^{\mathrm{v}}, \mathrm{X}^{0}),\quad \mathbf{C} = (0.09079, 0.08991, -0.08953),\quad \mathbf{u} = (5.7, 1.6, 0.9)$$

While the utility of the X_{VIS} term can strongly be enhanced (to 13) with the introduction of orthogonal terms, the utility of the other descriptors cannot be enhanced. The strong utility value of the orthogonalized term is in keeping with the fact that the orthogonal term encodes most of the description.

The modeling of the elutropic value, EV, of $n = 20$ solvents by M, ϵ, and a_{w} is

$$Q(M) = 3.1, F(M) = 10.8; Q(\epsilon) = 5.44, F(\epsilon) = 29; \text{and } Q(a_{\mathrm{w}}) = 3.5, F(a_{\mathrm{w}}) = 11.8$$

This description improves quickly with the use of molecular connectivity indices of a set of eq 1 (those of eq 2 are of inferior quality), where the ${}^{1}\chi^{\mathrm{v}}$ single index is a leading index

$$\{{}^{1}\chi^{\mathrm{v}}\}: Q = 8.19, F = 65.6, r = 0.886, s = 0.11$$

$$\{D^{\mathrm{v}}, {}^{1}\chi, {}^{1}\chi^{\mathrm{v}}\}: Q = 8.79, F = 25.2, r = 0.909, s = 0.10$$

As in our set of compounds, there are only two alcohols; it proves better to assume even for them $a_{\mathrm{w}} = a_{\epsilon} = \epsilon/15$, and in this case: $Q(a_{\epsilon}) = 2.41$ and $F(a_{\epsilon}) = 11.8$. Running then the corresponding supraindices of eq 9, where a_{w} is replaced by a_{ϵ}, yields the following single and multiple optimal descriptions, where, again, the single supraindex is a leading index

$$\{a_{\epsilon}{}^{1}\chi^{\mathrm{v}}\}: Q = 8.53, F = 71.7, r = 0.893, s = 0.11$$

$$\{a_{\epsilon}D^{\mathrm{v}}, a_{\epsilon}{}^{1}\chi, a_{\epsilon}{}^{1}\chi^{\mathrm{v}}\}: Q = 9.93, F = 32.2, r = 0.926, s = 0.09$$

This single supraindex description can be improved, with the introduction of the semiempirical term of eq 18, to $Q = 9.20$, $F = 82.8$, $r = 0.906$, $s = 0.1$, and $\langle u \rangle = 9.4$

$$\mathrm{X}_{\mathrm{EV}} = (a_{\epsilon}{}^{1}\chi^{\mathrm{v}} + 0.3)^{0.2} \qquad (18)$$

With a linear combination of this term plus a supraindex we obtain the following optimal description

$$\{X_{\mathrm{EV}}, a_{\epsilon}{}^{1}\chi\}: Q = 10.5, F = 54.3, r = 0.930, s = 0.09, \langle u \rangle = 7.5$$

The following vectors can be used to model the EV values

$$\mathbf{X} = (X_{\mathrm{EV}}, a_{\epsilon}{}^{1}\chi, \mathrm{X}^{0}),\quad \mathbf{C} = (-3.82941, 0.09413, 4.59876),\quad \mathbf{u} = (9.45, 2.33, 10.8)$$

If the two strongest outliers, 112tCltFEthane and pentane are excluded, the two index description $\{X_{\mathrm{EV}}, a_{\epsilon}{}^{1}\chi\}$ can be improved to $Q = 13.8$, $F = 77.4$, and $r = 0.955$.

TABLE 5: Experimental (exp) and Calculated (clc) UV Cutoff Values of $n = 27$ Solvents[a]

solvent	UV_{exp}	UV_{clc}		UV_{exp}	UV_{clc}
acetone*	330	269.9	1,4-dioxane	215	217.2
acetonitrile*	190	247.5	ether	215	222.2
benzene*	280	213.7	ethyl acetate	260	250.1
1-butanol	215	219.8	ethyl alcohol	210	219.7
2-butanone	330	307.8	heptane	200	228.0
butyl acetate	254	247.8	hexane*	200	242.1
CS_2*	380	299.2	2-methoxyethanol	220	228.3
CCl_4*	263	254.1	methyl alcohol	205	215.2
Cl-benzene*	287	250.6	4-M-2-pentanone	334	307.4
1Cl-butane	225	216.5	DMSO	268	250.9
$CHCl_3$	245	225.4	nitromethane*	380	328.9
cyclohexane	200	220.9	pentane*	200	264.0
cyclopentane	200	229.9	1-propanol	210	219.8
1,2-dCl-benzene	295	310.6	2-propanol	210	227.6
1,2-dCl-ethane	225	212.2	pyridine*	305	325.2
dCl-methane	235	218.4	tt-hydrofuran*	215	247.4
N,*N*-dM-acetamd	268	331.0	toluene*	285	217.3
N,*N*-dM-formamd	268	303.1	1,1,2tCltFEthane	230	236.4
			2,2,4-tM-pentane	215	240.5

[a] The 12 solvents not included in the fine simulation are denoted by an asterisk.

The simulation of $n = 37$ UV cutoff values starts very badly and hardly goes on with $\{D^{\mathrm{v}}\}$, $Q = 0.005$, $F = 2.51$, $r = 0.259$, a result that seems to mimic the bad results that can be obtained with M, ϵ, and a_{w}: $Q(M) = 0.002$, $F(M) = 0.26$; $Q(\epsilon) = 0.003$, $F(\epsilon) = 0.73$; $Q(a_{\mathrm{w}}) = 0.0002$, $F(a_{\mathrm{w}}) = 0.004$. Even introduction of the following type of semiempirical supraindices,

$$\{(\epsilon/a_{\mathrm{w}})(\chi,\chi^{\mathrm{v}})\} = \{\epsilon D/a_{\mathrm{w}}, \epsilon D^{\mathrm{v}}/a_{\mathrm{w}}, \epsilon^{0}\chi/a_{\mathrm{w}}, \epsilon^{0}\chi^{\mathrm{v}}/a_{\mathrm{w}}, \epsilon^{1}\chi/a_{\mathrm{w}}, \epsilon^{1}\chi^{\mathrm{v}}/a_{\mathrm{w}}, \chi_{\mathrm{t}}a_{\mathrm{w}}/\epsilon, \chi_{\mathrm{t}}{}^{\mathrm{v}}a_{\mathrm{w}}/\epsilon\} \qquad (19)$$

does not improve the modeling in a satisfactory way; in fact, the best single supraindex $\{\epsilon D^{\mathrm{v}}/a_{\mathrm{w}}\}$ rates only $Q = 0.01$, $F = 11.9$, $r = 0.504$. A closer look at the set of compounds lets us notice that (i) leaving out 12 nonalcoholic solvents (with asterisks in Table 5) and (ii) introducting the a_{OH} parameter already used along the modeling of viscosity, the modeling of the remaining $n = 25$ solvents improves consistently at the level of supraindices of eq 19, with a_{OH} replacing a_{w}. With $n = 25$, M, ϵ, and a_{OH} rate $Q(M) = 0.007$, $F(M) = 1.5$; $Q(\epsilon) = 0.007$, $F(\epsilon) = 1.90$; $Q(a_{\mathrm{OH}}) = 0.006$, $F(a_{\mathrm{OH}}) = 1.41$. Here, the best normal χ index rates rather badly, while a single supraindex guarantees a satisfactory model, linear combinations of supraindices bringing no further improvement,

$$\{D^{\mathrm{v}}\}: Q = 0.009, F = 2.90, r = 0.33, s = 36.8$$

$$\{\epsilon D/a_{\mathrm{OH}}\}: Q = 0.063, F = 139.2, r = 0.926, s = 14.7$$

A trial-and-error procedure discovers the brilliant semiempirical term of eq 20 centered around D^{v}, a_{OH}, and ϵ indices. This term, which rates $Q = 0.104$, $F = 380$, $r = 0.971$, $s = 9.33$, $\langle u \rangle = 41.9$, will be used to model the 25 UV values

$$X_{\mathrm{UV}} = \epsilon[(D^{\mathrm{v}})^{0.7} - 0.05a_{\mathrm{OH}}]/(a_{\mathrm{OH}})^{1.5} \qquad (20)$$

Linear combinations of this term with the supraindices of eq 19, with a_{OH} instead of a_{w}, show interesting improvements in Q, r, and s, in fact

$$\{X_{\mathrm{UV}}, \epsilon^{1}\chi/a_{\mathrm{OH}}, \epsilon^{1}\chi^{\mathrm{v}}/a_{\mathrm{OH}}\}: Q = 0.127, F = 190, r = 0.982, s = 7.7$$

The following vectors, based on the single term of eq 20, can be used to model the UV cutoff values.

$$\mathbf{X} = (X_{\mathrm{UV}}, X^0), \qquad \mathbf{C} = (1.26155, 192.494), \qquad \mathbf{u} = (19.5, 64.3)$$

Now let us reintroduce the 12 strong outliers, excluded in the previous modeling, and handle the $n = 37$ solvents, with a set of eq 19 and let us try to calculate their UV cutoff values with the following best linear combination, $\{\epsilon D/a_{\mathrm{w}}, \epsilon D^{\mathrm{v}}/a_{\mathrm{w}}, \epsilon^1\chi/a_{\mathrm{w}}, \chi_{\mathrm{t}}^{\mathrm{v}} a_{\mathrm{w}}/\epsilon\}$, which even with a poor utility and a not impressive predictive power with $Q = 0.019$, $F = 16.4$, and $r = 0.70$, is nevertheless susceptible to get interesting calculated 37 UV cutoff values, especially for the $n = 25$ solvents fitted in the previous modeling, as can be seen in Table 5.

Modeling the polarity of 36 solvents with molecular connectivity indices, Kier[54] concluded that the ${}^1\chi^{\mathrm{v}}/f$ index, where f describes the number of discrete, isolated functional groups, was the best descriptor for this property. More recently, by the aid of an empirical index based on molecular polarizabilities, an excellent description of some properties, related to water solubilities of alkanes and alcohols has been achieved.[55] It is interesting to notice that also for the modeling of the dipole moment of $n = 35$ solvent of Table 2, the ${}^1\chi^{\mathrm{v}}$ index is the best index, even if its ratings are rather poor, while M, ϵ, and a_{w} rate $Q(M) = 0.18$, $F(M) = 1.9$; $Q(\epsilon) = 0.74$, $F(\epsilon) = 32.5$; and $Q(a_{\mathrm{w}}) = 0.6$, $F(a_{\mathrm{w}}) = 24.7$. The best molecular connectivity descriptor, worse than M, is $\{{}^1\chi^{\mathrm{v}}\}$ with $Q = 0.17$, $F = 1.8$, $r = 0.23$, $s = 1.3$, while with $\{{}^1\chi^{\mathrm{v}}/f\}$ index, things improve only a little bit, with $Q = 0.21$ and $F = 1.83$. Introduction and use of the following set of semiempirical descriptors

$$\{(\epsilon)(\chi, \chi^{\mathrm{v}})\} = \{\epsilon D, \epsilon D^{\mathrm{v}}, \epsilon^0\chi, \epsilon^0\chi^{\mathrm{v}}, \epsilon^1\chi, \epsilon^1\chi^{\mathrm{v}}, \chi_{\mathrm{t}}/\epsilon, \chi_{\mathrm{t}}^{\mathrm{v}}/\epsilon\} \tag{21}$$

improves consistently the description with a single index and a two-index combination

$$\{\epsilon^0\chi\}: Q = 1.06, F = 67.5, r = 0.82, s = 0.77$$

$$\{\epsilon^0\chi, \chi_{\mathrm{t}}/\epsilon\}: Q = 1.35, F = 55.0, r = 0.88, s = 0.65$$

After that the description ameliorates rather slowly, but the decisive improvement is obtained with the following semiempirical term, which rates $Q = 2.12$, $F = 272.0$, $r = 0.94$, $s = 0.44$, $\langle u \rangle = 9.7$

$$X_\mu = [(\epsilon D^{\mathrm{v}} - 1.9D)/(2.5D^{\mathrm{v}} - \epsilon\chi_{\mathrm{t}})]^{0.45} \tag{22}$$

This single term can be used to model the dipole moments of 35 solvents, and its vectors are

$$\mathrm{X} = (X_\mu, X^0), \qquad \mathbf{C} = (1.43421, -0.46761), \qquad \mathbf{u} = (16.5, 2.9)$$

The description of this property can further be improved at the Q, r, and s level with the following two-index combination: $\{X_\mu, {}^1\chi^{\mathrm{v}}\}$ with $Q = 2.24$, $F = 151$, $r = 0.951$, $s = 0.42$, $\langle u \rangle = 7.4$ (δ^{z}-based indices rate rather poorly, at every level).

The modeling of the magnetic susceptibility MS of 19 solvents of Table 2 with M, ϵ, and a_{w}, [$Q(M) = 10.5$, $F(M) = 22.1$; $Q(\epsilon) = 6.8$, $F(\epsilon) = 9.4$; $Q(a_{\mathrm{w}}) = 8.5$, $F(a_{\mathrm{w}}) = 14.5$], when compared with the rating of the best $\{{}^0\chi^{\mathrm{v}}\}$ descriptor, cannot be underestimated. The following result, $Q({}^0\chi^{\mathrm{v}}) = 8.4$, $F({}^0\chi^{\mathrm{v}}) = 14.4$, $r({}^0\chi^{\mathrm{v}}) = 0.62$, $s({}^0\chi^{\mathrm{v}}) = 0.08$ (${}^1\chi^{\mathrm{v}}$ shows nearly the same quality), indicates that single molecular connectivity indices do not offer a better modeling than empirical indices, and δ^{z}-based χ indices fare even worse. The following linear combination of two χ indices shows promising results, $\{{}^1\chi, \chi_{\mathrm{t}}^{\mathrm{v}}\}$, $Q = 13.6$, $F = 18.8$, $r = 0.84$, $s = 0.08$, but further improvement, with the introduction of more indices, is impossible to achieve. Introduction of the following set of semiempirical molecular connectivity indices,

$$\{(a_{\mathrm{w}}{}^{-1})(\chi, \chi^{\mathrm{v}})\} = \{D/a_{\mathrm{w}}, D^{\mathrm{v}}/a_{\mathrm{w}}, {}^0\chi/a_{\mathrm{w}}, {}^0\chi^{\mathrm{v}}/a_{\mathrm{w}}, 1\chi/a_{\mathrm{w}}, {}^1\chi^{\mathrm{v}}/a_{\mathrm{w}}, \chi_{\mathrm{t}} a_{\mathrm{w}}, \chi_{\mathrm{t}}^{\mathrm{v}} a_{\mathrm{w}}\} \tag{23}$$

let us detect somewhat better single and double descriptors with

$$\{{}^0\chi^{\mathrm{v}}/a_{\mathrm{w}}\}: Q = 10.9, F = 24, r = 0.76, s = 0.07$$

$$\{D/a_{\mathrm{w}}, \chi_{\mathrm{t}}^{\mathrm{v}} a_{\mathrm{w}}\}: Q = 15.6, F = 24.7, r = 0.87, s = 0.06$$

But even here further improvement cannot be achieved. The quest for a semiempirical molecular connectivity term ends up with the following very interesting, but rather convoluted term

$$X_{\mathrm{MS}} = [{}^1\chi M^{0.7} + 3(D^{\mathrm{v}})^{0.7}]/[D^{\mathrm{v}} + 2({}^0\chi^{\mathrm{v}})^{1.5} - 1.5({}^0\chi)^{1.4}] \tag{24}$$

This term rates $Q = 26.6$, $F = 143$, $r = 0.945$, $s = 0.04$, $\langle u \rangle = 7.7$. The following vectors can be used to simulate this property

$$\mathbf{X} = (X_{\mathrm{MS}}, X^0), \qquad \mathbf{C} = (0.20043, -0.24624), \qquad \mathbf{u} = (12, 3.4)$$

This term is a dead-end descriptor, and it cannot be used in any combination with other indices to further improve the modeling.

Many of our semiempirical modelings are based either on ϵ or on approximate ϵ- and H-bond-related parameters (a_{w}, a_{OH}, or a_ϵ), which are not exceptional descriptors of the dielectric constant ϵ. The search for a description for this property ($n = 63$) with molecular connectivity δ^{z}- and δ^{v}-based indices of sets of eqs 1 and 2 lands on the following best but bad description $\{{}^0\chi^{\mathrm{v}}\} \approx \{{}^1\chi^{\mathrm{v}}\}$: $Q = 0.025$, $F = 13.5$, $r = 0.42$, which strongly mimics the descriptive power of M, with $Q = 0.02$, $F = 8.2$. Interesting enough is the fact that ${}^1\chi^{\mathrm{v}}$ is the best molecular connectivity descriptor for dipole moment, susceptibility, and dielectric constant (and polarity, see refs 54 and 55). The trial-and-error search for an optimal semiempirical term for this property discovers the following dead-end term

$$X_\epsilon = (a_{\mathrm{w}}{}^0\chi + \chi_{\mathrm{t}}^{\mathrm{v}})^{0.75}/(M + 10^{4.0}\chi^{\mathrm{v}})^{0.6} \tag{25}$$

with $Q = 0.12$, $F = 321$, $r = 0.92$, $s = 7.7$, $\langle u \rangle = 13.8$.

Conclusions

Use of semiempirical methods is not new as it is well-known in quantum chemistry, or more generally in computational chemistry, e.g., in molecular mechanics (MM) and dynamics (MD). Even the history of molecular connectivity modeling is labeled by the use of empirical parameters linearly combined with χ indices (see ref 13 and references therein).

We have introduced here, for the first time, semiempirical molecular connectivity terms, i.e., special composition of molecular connectivity indices based both on topological and on empirical parameters. While the topological parameters used here are the well-known molecular connectivity indices of eq

1, the chosen empirical parameters are the dielectric constant, usually used in molecular dynamics studies to mimic the solvent, the molar masses, and novel 'ad hoc' empirical parameters, a_w, a_{OH}, and a_ϵ, based on dielectric constant and hydrogen bond considerations. It should be remembered that the dielectric constant is also important in Raman studies, where scattering and absorption contributions in colloids are dependent on ϵ of the dispersing medium.[56] The introduction of these empirical parameters has been done to overcome the inherent limits of χ indices, which tend to underestimate noncovalent interactions between compounds, in different states of aggregation. This strategy to develop semiempirical molecular connectivity terms appears promising in QSPR studies (quantitative structure−property relationships); it remains to be investigated how useful such terms will be in structure activity studies. The wealth of semiempirical terms found all along this study shows that (i) these semiempirical terms are not always elementary in their form, even if the formalism to derive them is rather straightforward, (ii) one cannot be absolutely confident that a better term, or a better empirical parameter, is hiding somewhere else, and (iii) many times the found term is a dead-end term, allowing no further improvement with multiple regressions. This last fact calls for descriptions with as few parameters as possible, a fact positive in its own; in fact, melting points, density, flashpoints, UV cutoff, dipole moments, magnetic susceptibility, and the same dielectric constant have been modeled with just one semiempirical term, while viscosity and elutropic values have been modeled with a single semiempirical term plus a semiempirical supraindex. Only the more than 60 values of boiling points and refractive indices have been modeled with 5 descriptors each. But let us look in more detail at the achieved results: the modeling of the 11 properties hints to the possibility of finding an optimal semiempirical term for each property, once the right empirical parameter has been detected. The intricate modeling of the melting temperatures can be enhanced using a 'double sieve' modeling strategy, possibly grounded on well-known noncovalent interactions, but for the moment, due to the lack of quantitative information about this kind of interactions, the 'double sieve' mechanism can only be imprecisely inferred. Interesting is the possibility to model with the same kind of semiempirical term, $X_{MP,BP}$, both for melting and for boiling points, which seems rather sound if the problem of a general theory about significant temperatures has to be raised. This last common term for melting and boiling points is based on nonvalence supraindices, which are contributed by the sum-δ, D, and the total connectivity index, χ_t; practically two bulk supraindices determine this kind of modeling. Boiling and melting points terms, X_{BP} and X_{MP}, individually, are based on supramolecular connectivity indices alone, attesting the importance of noncovalent and hydrogen bond interactions. The refractive index, RI, is the only property described by a pure molecular connectivity term, X_{RI}, based on both total molecular connectivity indices, χ_t^v and χ_t. The term for density, X_d, requires the consonant help of molar masses.The elutropic value term, X_{EV}, is constructed with supramolecular indices based on the 'ad hoc' empirical parameter $a_\epsilon = \epsilon/15$, with no special value for hydrogen bond; the dipolar moment term, X_μ, is based on the dielectric constant, which seems rather sound. The remaining other terms need the help of more than one empirical parameter; the flashpoint term, X_{FP}, is a function of M, ϵ, and supramolecular connectivity indices based on a_w; X_ϵ term is function of M and of supraindices grounded on a_w; the term for viscosity, X_{VIS}, is function of supraindices grounded both on $a_{OH} = 2 + \epsilon/15$, and on M (i.e., it is empirically influenced by bulk, noncovalent, and hydrogen bond characteristics); and, finally, the term for UV cutoff, X_{UV}, is a function of the dielectric constant and of the a_{OH} parameter alone. Generally, δ^v-based valence indices χ^v of eq 1 seem more appropriate, with the exception of melting and flashpoint values, to construct the molecular connectivity terms, even if both kinds of molecular connectivity indices alone, χ^v and χ^z, offer a bad description of the given properties. Equations for the different properties are very much in the spirit of QSPR (quantitative structure−property relationships); they tell us how given properties depend on the interplay of empirical parameters and topological descriptors. The three main goals of this paper, expounded at the end of the Introduction, have been for a good extent achieved, and from the given study, a rather clear picture of the general characteristics for the missing term can be deduced. It can, in fact, be said that this term should include (i) information on hydrogen bonding in a rather dynamical way (see the interplay between a_{OH}, a_ϵ, and a_w), (ii) information on noncovalent interactions and polarity, possibly in a highly detailed way, especially for the melting points, and (iii) more direct information on molar masses. A solution not easy to find.

Received wisdom has it that for every complex problem, there is a simple solution, even if the Italian writer Umberto Eco added to it "and it's wrong".

Acknowledgment. I would like to thank Prof. L. B. Kier (Virginia University) for some copies of his works and some discussions about polarity, both reviewers for hints to improve the paper, and especially one of them who overtook even the painful task of checking the names of every compound. To Professor M. Randič (Drake University, Iowa) I owe interesting remarks about the recent works of A. R. Katritzky et al. Finally, I would like to thank the Editor G. W. Milne for sending me a precious copy of the last number of *J. Chem. Inf. Comput. Sci.* MURST, the ministry for University, Scientific, and Technological Research, is gratefully acknowledged for financial support.

References and Notes

(1) Pogliani, L. *J. Phys. Chem.* **1995**, *99*, 925−937.
(2) Pogliani, L. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1082−1096.
(3) Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065−18077.
(4) Pogliani, L. *Croat. Chem. Acta* **1997**, *70*, 803−817.
(5) Pogliani, L. *Med. Chem. Res.* **1997**, *7*, 380−393; *Amino Acids* **1997**, *13*, 237−255.
(6) Pogliani, L. *J. Mol. Struct. (THEOCHEM)*, in press.
(7) Pogliani, L. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 104−111.
(8) Randič, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(9) Kier, L. B.; Hall, L. H.; Murray, W. J.; Randič, M. *J. Pharm. Sci.* **1975**, *64*, 1971−1974.
(10) Hall, L. H.; Kier, L. B.; Murray, W. J. *J. Pharm. Sci.* **1975**, *64*, 1974−1977.
(11) Murray, J. M.; Hall, L. H.; Kier, L. B. *J. Pharm. Sci.* **1975**, *64*, 1978−1981.
(12) Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1976**, *65*, 1806−1809.
(13) Kier,.B.; Hall, H. H. *Molecular Connectivity in Structure−Activity Analysis*; Wiley: New York, 1986.
(14) Balaban, A. T., Ed. *Chemical Applications of Graph Theory*; Academic Press: London, 1976.
(15) Kier, L. B.; Hall, L. H. *J. Pharm. Sci* **1981**, *70*, 583−589.
(16) Trinajstič, N. *Chemical graph theory*; CRC Press: Boca Raton, FL, 1983 (second ed., 1992).
(17) Randič, M.; Hansen, P. J.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60−68.
(18) Needham, D. E.; Wei, I.-C.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 4186−4194.
(19) Rouvray, D. H. *J Mol Struct (THEOCHEM)* **1989**, *185*, 187−201.
(20) Randič, M. *New J. Chem.* **1991**, *15*, 517−525.
(21) Randič, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.
(22) Randič, M. *Croat. Chim. Acta* **1991**, *64*, 43−54.

(23) Basak, S. C.; Niemi, G. J.; Veith, G. D. *J. Math. Chem* **1991**, *7*, 243−272.

(24) Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23−28.

(25) Maier, B. J. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 87−90.

(26) Mihalič, Z.; Trinajstič, N. *J. Chem. Educ.* **1992**, *69*, 701−712.

(27) Mihalič, Z.; Nikolič, S.; Trinajstič, N. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28−37.

(28) Pogliani, L. *J. Pharm. Sci.* **1992**, *81*, 334−336.

(29) Pogliani, L. *J. Pharm. Sci.* **1992**, *81*, 967−969.

(30) Stanton, D. T.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 109−115.

(31) Kier, L. B,; Hall, L. H.; Frazer, J. W. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 143−147.

(32) Hall, L. H.; Kier, L. B.; Frazer, J. W. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 148−152.

(33) Pogliani, L. *Comput. Chem.* **1993**, *17*, 283−286.

(34) Pogliani, L. *J. Phys. Chem.* **1993**, *97*, 6731−6736.

(35) Randič, M. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1994**, *21*, 215−225.

(36) Basak, S. C.; Grunwald, G. D. *SAR & QSAR Environ. Res.* **1994**, *2*, 289−307.

(37) Pogliani, L. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 801−804.

(38) Galvez, J.; Garcia-Domenech, R.; de Julian-Ortiz, J. V. ; Soler, R. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272−284.

(39) Galvez, J.; Garcia-Domenech, R.; de Gregorio Alapont C.; de Julian-Ortiz, J. V.; Popa, L. *J. Mol. Graph.* **1996**, *14*, 272−276.

(40) Garcia-Domenech, R.; de Gregorio Alapont, C.; de Julian-Ortiz, J. V.; Galvez, J. *Bioorg., Med. Chem. Lett.* **1997**, *7*, 567−572.

(41) Galvez, J. *J. Mol. Struct.(THEOCHEM)* **1998**, *429*, 255−264.

(42) Kuanar, M.; Mishra, B. K *Bull. Chem. Soc. Jpn.* **1998**, *71*, 191−198.

(43) *Bruker Almanac*; Bruker Scientific Instruments: Rheinstetten, Germany, 1986; pp 106−107.

(44) *Aldrich Solvents*; Aldrich Chemical Co., Inc.: Milwaukee, WI, 1995; p 9.

(45) Lide, R. D., Ed.-in-Chief. *CRC Handbook of Chemistry and Physics*, 74th ed.; CRC: Boca Raton, FL, 1993−1994; pp 6-148−6-155.

(46) Myers, R. T. *J. Phys. Chem.* **1979**, *83*, 294−296.

(47) Rich, R. L. *Bull. Chem. Soc. Jpn.* **1993**, *66*, 1065−1078.

(48) Hansen, P. J.; Jurs, P. C. *Anal. Chem.* **1987**, *59*, 2322−2327.

(49) Balaban, A. T.; Basak, S. C.; Colburn, T.: Grunwald, G. D. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118−1121.

(50) Carlton, T. S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 158−164.

(51) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28−41.

(52) Katritzky, A. R.; Sild, S.; Lobanov, V. ; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 300−304.

(53) Katritzky, A. R.; Mu, L.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 293−299.

(54) Kier, L. B. *J. Pharm. Sci.* **1981**, *70*, 930−932.

(55) Chenzhong, C.; Zhiliang, L. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1−7.

(56) Creighton, J. A. In *Surface Enhanced Raman Spectroscopy*; Chang, R. K., Furtak, T. E., Eds.; Plenum: New York, 1982; pp 315−338.